



# STIC Search Report

## Biotech-Chem Library

### STIC Database Tracking Number

TO: Georgia L Helmer  
Location: 9d14 / 9e12  
Friday, November 07, 2003  
Art Unit: 1638  
Phone: 308-7023  
Serial Number: 09 / 622500

From: Jan Delaval  
Location: Biotech-Chem Library  
CM1-1E07  
Phone: 308-4498

jan.delaval@uspto.gov

### Search Notes

Edward Hart submitted your jobs to queue on November 7, 2003.

I have printed the results for you. Any questions should be directed to Edward at your convenience next week.

*pending  
removed*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:50:29 : Search time 1314 Seconds

(without alignments)  
4883.09: Million cell updates/sec

Title: US-09-622-500b-3

Sequence: 264  
1 atgacaaagtggtgagatcaaa.....gccctgtgcccgcacgaatga 264

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Word size: 3

Search: 22781392 seqs, 1258238056 residues

45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: em\_estba:  
2: em\_esthm:  
3: em\_estim:  
4: em\_estnu:  
5: em\_estov:  
6: em\_estpl:  
7: em\_estro:  
8: em\_hic:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hic:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estun:  
16: em\_estom:  
17: em\_gss\_hum:  
18: em\_gss\_inv:  
19: em\_gss\_pln:  
20: em\_gss\_vrt:  
21: em\_gss\_fun:  
22: em\_gss\_mam:  
23: em\_gss\_mus:  
24: em\_gss\_pro:  
25: em\_gss\_rtd:  
26: em\_gss\_phg:  
27: em\_gss\_vrl:  
28: gb\_gss1:  
29: gb\_gss2:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	8.0	474	13	BQ280354
2	19	7.2	136	28	A2601419
3	19	7.2	497	14	C8220913
4	19	7.2	656	10	BG705752

5	19	7.2	681	10	B2370873
6	19	7.2	727	28	AQ544745
7	19	7.2	766	29	B2571110
8	19	7.2	889	13	BUR57584
9	19	7.2	942	12	E1078258
10	19	7.2	970	29	CNSC0731
11	19	7.2	1112	29	B2577558
12	19	7.2	1122	29	A4489437
13	19	7.2	140	9	A4726470
14	19	7.2	140	9	AV940990
15	19	7.2	140	9	AV940990
16	19	7.2	140	9	AV940990
17	19	7.2	140	9	AV940990
18	19	7.2	140	9	AV940990
19	19	7.2	140	9	AV940990
20	19	7.2	140	9	AV940990
21	19	7.2	140	9	AV940990
22	19	7.2	140	9	AV940990
23	19	7.2	140	9	AV940990
24	19	7.2	140	9	AV940990
25	19	7.2	140	9	AV940990
26	19	7.2	140	9	AV940990
27	19	7.2	140	9	AV940990
28	19	7.2	140	9	AV940990
29	19	7.2	140	9	AV940990
30	19	7.2	140	9	AV940990
31	19	7.2	140	9	AV940990
32	19	7.2	140	9	AV940990
33	19	7.2	140	9	AV940990
34	19	7.2	140	9	AV940990
35	19	7.2	140	9	AV940990
36	19	7.2	140	9	AV940990
37	19	7.2	140	9	AV940990
38	19	7.2	140	9	AV940990
39	19	7.2	140	9	AV940990
40	19	7.2	140	9	AV940990
41	19	7.2	140	9	AV940990
42	19	7.2	140	9	AV940990
43	19	7.2	140	9	AV940990
44	19	7.2	140	9	AV940990
45	19	7.2	140	9	AV940990

## ALIGNMENTS

RESULT 1  
BQ280354  
LOCUS  
DEFINITION  
OMNI2E10 Canine Brain cDNA library Caris familiaris CPNA 5' similar  
to Homo sapiens retinoid acid repressible protein, mRNA sequence.

ACCESSION  
BQ280354.1 GI:20520360  
VERSION  
BQ280354.1 GI:20520360  
KEYWORDS  
EST

SOURCE  
ORGANISM  
Caris familiaris (dog)  
Tissue: Brain

REFERENCE  
AUTHORS  
Roberts, M.C., Hendrickson, J.A., Hoffmann, D.E., Flickinger, G.H.,  
University of Minnesota Canine Brain EST Project

COMMENT  
Contact: Mickelson, J.R.  
Veterinary Pathobiology  
University of Minnesota  
1988 Petch Avenue, University of Minnesota, St. Paul, MN 55106, USA  
Tel: 612 624 1246  
Fax: 612 625 0204  
Email: mickel@tc.umn.edu  
Seq primer: M13 Reverse  
Location/Qualifiers  
1. 474

FEATURES  
SOURCE

```

/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/sex="Male"
/clone_lib="Canine Brain cDNA Library"
/notes="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site: 1;
Note: Site 2: SalI; Tissue was taken from the frontal,
occipital, temporal and parietal lobes, olfactory bulb,
hippocampus, cerebellum, thalamus, hypothalamus, midbrain
, pons, and medulla."
BASE COUNT      121 a      115 c      167 g      60 t      11 others
ORIGIN

Query Match      8.0%; Score 21; DB 13; Length 474;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

146 TGCACGAGGCGCCAGCCGCT 166
|||||
61 TGCACGAGGCGCCAGCCGCT 61

RESULT 2
AZ601419      306 bp      DNA      linear      GSS 13-DEC-2000
LOCUS      1M0419L09R Mouse 10kb plasmid UUCGM library Mus musculus genomic
DEFINITION      clone UUCGM0419L09 R, genomic survey sequence.
ACCESSION      AZ601419
VERSION      AZ601419.1 GI:11723609
KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 306)
Dunn,D., Hoyagi,A., Barber,X., Seacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
Unpublished
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5636
Fax: 801 585 7177
Email: dcunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0419 row: L column: 09
Seq primer: CACACGAGAAACAGCATGACCC
Class: plasmid ends
High quality sequence scop: 306.
Location/Qualifiers
1..306
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCGM0419L09"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42rv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/charters/). The DNA
was hydrodynamically sheared by repeated passage through a
3.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were

```

```

ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi|473211419|AF19372.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
cloned vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      84 a      66 c      96 g      60 t
ORIGIN

Query Match      7.2%; Score 19; DB 28; Length 306;
Best Local Similarity 100.0%; Pred. No. 2,2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 CAGCCGAGCCAGCCGAG 36
|||||
138 CAGCCGAGCCAGCCGAG 156

RESULT 3
CB220913      497 bp      mRNA      linear      EST 10-FEB-2003
LOCUS      1AB027H07 Bos taurus Abomasum #1 library Bos taurus cDNA, mRNA
DEFINITION      sequence.
ACCESSION      CB220913
VERSION      CB220913.1 GI:26291427
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 497)
Hansen,C., Fu,A., Meng,Y., Li,C., Okunev,E., Sersen,C.W., Gordon
,P.V.K. and Moore,S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
JOURNAL
Unpublished
Contact: Dr. Stephen Moore
Reef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P6, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ua.bertha.ca
Insert Length: 497 Std Error: 0.00
POLYVA=NC.
Location/Qualifiers
1..497
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Smooth muscle"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-Biomer"
/clone_lib="Bos taurus Abomasum #1 library"
/notes="Organ: Abomasum; Vector: Uni-Z24PKS; Site: 1; EPCR
I; Site 2: Xho I"
BASE COUNT      136 a      124 c      174 g      63 t
ORIGIN

Query Match      7.2%; Score 19; DB 14; Length 497;
Best Local Similarity 100.0%; Pred. No. 2,5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

146 TGCACGAGGCGCCAGCCGCG 164
|||||
75 TGCACGAGGCGCCAGCCGCG 93

```

RESULT 4  
 BG705752 656 bp mRNA linear EST 07-MAY-2001  
 LOCUS 60266891P1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4791574 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG705752  
 VERSION BG705752.1 GI:13980409  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.  
 AUTHORS 1 (bases 1 to 656)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael C. Brownstein (NHGRI), Shizaki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHAM10668 row: 6 column: 23  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1..656  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4791574"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_96"  
 /note="Organ: brain; Vector: pBluescript (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI; XhoI; Gcgag  
 ); 01190-dt primed using primer 5'-TTTCTTTTCTTTCTTTCTTT-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to RCT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 144 a 200 c 206 g 106 t  
 ORIGIN  
 Query Match 7.2% Score 19; DB 10; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 2,66+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 CAAGGTGAGATCAAGCCC 24  
 DB 469 CAAGGTGAGATCAAGCCC 487  
 RESULT 5  
 BE370873 681 bp mRNA linear EST 21-JUN-2003  
 LOCUS 601219090P1 NCI\_CGAP\_1029 Mus musculus cDNA clone IMAGE:3588046 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE370873  
 VERSION BE370873.1 GI:9316236  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHAM8752 row: 3 column: 23  
 High quality sequence stop: 54.  
 High quality sequence start: 544.  
 Location/Qualifiers  
 1..681  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C280H II"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3588046"  
 /tissue\_type="spontaneous tumor, metastatic to mammary,  
 stem cell origin."  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_1029"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally; Primer: 01190 dt  
 Library constructed by Life Technologies, Investigator  
 providing samples: Gilbert Smith, NIH"  
 BASE COUNT 149 a 163 c 235 g 194 t  
 ORIGIN  
 Query Match 7.2% Score 19; DB 17; Length 681;  
 Best Local Similarity 100.0%; Pred. No. 2,76+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 134 TGAAGATCAACCTGGACCA 152  
 DB 61 TGAAGATCAACCTGGACCA 79  
 RESULT 6  
 A0544745 727 bp DNA linear GSS 28-MAY-1999  
 LOCUS A0544745  
 DEFINITION C17B1-E1-261G7.TF C17B1-E1 Homo sapiens genomic clone 2611G7,  
 genomic survey sequence.  
 ACCESSION A0544745  
 VERSION A0544745.1 GI:4903556  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE Zhao S., Adams, W.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and  
 Venter, J. C.  
 TITLE Use of BAC End Sequences from Ca-Tech Libraries for Sequence-Ready  
 Map Building  
 JOURNAL Unpublished  
 COMMENT Other GSSs: C17B1-E1-261G7.TF  
 Contact: Shiyang Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9722 Medical Center Dr., Rockville, MD 20855  
 Tel: 301 838 0200  
 Fax: 301 838 0258  
 Email: [hbe@tigr.org](mailto:hbe@tigr.org)  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
 Seq primer: M13-21  
 Class: BAC ends.  
 Location/Qualifiers



```

source
    .: 727
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /clone="2611G7"
    /sex="male"
    /cell_type="sperm"
    /clone_id="CITB1-E"
    /note="Vector: pBluescript; Site_1: EcoRI; Site_2: EcoRI; Caltech Human BAC Library D"

BASE COUNT      210 a      146 c      186 g      185 t
ORIGIN

Query Match          7.2% Score 19; DB 28; Length 727;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      20 AGCCGAGCCAGGCCCGATT 38
        |||||
Db       557 AGCCGAGCCAGGCCCGATT 575

RESULT 7
LOCUS     BZ571110              766 bp    DNA             linear   GSS 17-DEC-2002
DEFINITION msh2_1733.x1 msh Pseudomonas aeruginosa genomic clone msh2_1733,
            genomic survey sequence.
VERSION     BZ571110
KEYWORDS    BZ571110.1 GI:27236171
SOURCE      GSS.
            Pseudomonas aeruginosa
            Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
            1 (bases 1 to 766)
REFERENCE   Spence,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
AUTHORS     Burns,J.L., Kaul,R. and Olsen,K.V.
TITLE       Whole-Genome-Sequence Variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL     J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 357145, Seattle, WA 98195-2145, USA
            Tel.: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES
    source
        .. 766
        /organism="Pseudomonas aeruginosa"
        /mol_type="genomic DNA"
        /strain="M5H"
        /db_xref="taxon:287"
        /clone="msh2_1733"
        /clone_id="msh"
        /note="Environmental isolate. Whole genomic shotgun
            library"

BASE COUNT      123 a      176 c      184 g      283 t
ORIGIN

Query Match          7.2% Score 19; DB 29; Length 766;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      126 CGTGGCTGGTGAGATCAC 144
        |||||||
Db       161 CGTGGCTGGTGAGATCAC 179

RESULT 8
LOCUS     BU857584/c              889 bp    mRNA             linear   EST 16-OCT-2002

```

```

DEFINITION
AGENCOURT 104735:7 NIH-MGC-001 Homo sapiens cDNA clone
IMAGE:6647448.5', mRNA sequence.
ACCESSION
U0857884
U0857884.1 GI:24042576
VERSION
EST
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Carnifera; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 889)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Robin Laboratory
cDNA Library Arrayed by: The I.V.A.G.E. Consortium (JUN-1)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.V.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Date: Dec2987 row: a column: 24
High quality sequence score: 596.
Location/Qualifiers
1..889
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6647448"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC 107"
/note="Organ: breast; Vector: pORF7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming;
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GTGACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library."
BASE COUNT
141 a 321 c 285 g 140 t 2 others
ORIGIN
Query Match: 7.2%, Score 19; DB 13; Length 889.
Best Local Similarity 100.0%; Pfad No. 2,5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Cy 18 CAAGCCGACGACGACCG 36
||| ||||| ||||| |||
Db 789 CAAGCCGACGACGACCG 771
RESULT 9
BI078258 942 bp mRNA linear EST 20-JUN-2000
LOCUS BI078258
DEFINITION 60287259CPI_MCI_CGAG_Mam2 Mus musculus cDNA clone IMAGE:5034263.5',
mRNA sequence.
ACCESSION BI078258
VERSION BI078258.1 GI:14396588
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Carnifera; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 942)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: WGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINT at:  
<http://image.lim.gov>  
 Plate: LLM11042 row: d column: 24  
 High quality sequence stop: 787.  
 Location/Qualifiers

yeast species: *Saccharomyces bayanus* var. *uvarum*. *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

## FEATURES

## source

1..942  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5034263"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 morchs"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Mam2"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

Location/Qualifiers  
 1..970  
 /organism="Kluyveromyces marxianus"  
 /mol\_type="genomic DNA"  
 /strain="CBS 712"  
 /variety="marxianus"  
 /db\_xref="taxon:4911"  
 /clone="AZ0AA014C12"  
 /clone\_lib="AZ0AA"  
 /note="end : 77"  
 <3..>370  
 /note="similar to *Saccharomyces cerevisiae* CBF YPH120 ;  
 similarity to RNA-binding proteins ;  
 1 putative frameshift(s)"  
 /evidence="not experimental"

## BASE COUNT

232 a 230 c 314 g 166 t

## BASE COUNT

278 a 196 c 251 t 1 others

## ORIGIN

Query Match 7.2%; Score 19; DB 12; Length 942;  
 Best Local Similarity 100.0%; Pred. No. 2,9e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## Query Match

7.2%; Score 19; DB 29; Length 970;  
 Best Local Similarity 100.0%; Pred. No. 2,9e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 GGTGAAGATCACCCTGGAC 150  
 |||||  
 10 GGTGAAGATCACCCTGGAC 28

QY 36 GTTACACACCCGACGCGC 54  
 |||||  
 878 GTTACACACCCGACGCGC 860

RESULT 10  
 CDS0733L/C 970 bp DNA linear GSS 06-JUL-2001  
 LOCUS 77 end of clone AZ0AA014C12 of library AZ0AA from strain CBS 712 of  
 DEFINITION *Kluyveromyces marxianus*, genomic survey sequence.  
 ACCESSION AL424423  
 VERSION AL424423.1 GI:12207617  
 KEYWORDS GSS.

RESULT 11  
 B2577558 1111 bp DNA linear GSS 17-DEC-2002  
 LOCUS B2577558  
 DEFINITION *Pseudomonas aeruginosa* genomic clone msh2\_5464,  
 genomic survey sequence.

ACCESSION AL424423.1 GI:12207617  
 VERSION AL424423.1  
 KEYWORDS GSS.  
 SOURCE *Kluyveromyces marxianus*  
 ORGANISM *Kluyveromyces marxianus*  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; *Kluyveromyces*;  
 1 (bases 1 to 970)

ACCESSION B2577558  
 VERSION B2577558  
 KEYWORDS GSS.  
 SOURCE *Pseudomonas aeruginosa*  
 ORGANISM *Pseudomonas aeruginosa*  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; *Pseudomonas*;  
 1 (bases 1 to 1111)

## REFERENCE

AUTHORS Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Boloitin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
 de-Montigny,J., Dujon B., Duren P., Lepligle,A., Llorente,B.,  
 Malpertuy,A., Neveuglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekaia,F., Toffano-Nicche,C., Wesolowski-Douvet,M.,  
 Wincker,P. and Weissenbach,C.  
 Genomic exploration of the hemiascomycetous yeasts : I. A set of  
 yeast species for molecular evolution studies

REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Ojser,M.V.  
 Whole-genome-sequence variation among multiple isolates of  
*Pseudomonas aeruginosa* library  
 J. Bacteriol., (2002) in press

## TITLE

Genomic exploration of the hemiascomycetous yeasts : I. A set of  
 yeast species for molecular evolution studies

## AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Ojser,M.V.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## COMMENT

Genome Center  
 Contact: Chris K. Raymond  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [ckraymond@u.washington.edu](mailto:ckraymond@u.washington.edu)  
 Class: shotgun.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## MEDLINE

1152876

## FEATURES

Location/Qualifiers  
 1..1111  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="msh2\_5464"  
 /clone\_lib="msh"  
 /note="Environmental isolate, whole genomic shotgun  
 library"

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Ojser,M.V.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## COMMENT

Genome Center  
 Contact: Chris K. Raymond  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [ckraymond@u.washington.edu](mailto:ckraymond@u.washington.edu)  
 Class: shotgun.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Ojser,M.V.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## COMMENT

Genome Center  
 Contact: Chris K. Raymond  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [ckraymond@u.washington.edu](mailto:ckraymond@u.washington.edu)  
 Class: shotgun.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Ojser,M.V.

## JOURNAL

FEBS Lett. 487 (1), 71-75 (2000)

## COMMENT

Genome Center  
 Contact: Chris K. Raymond  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [ckraymond@u.washington.edu](mailto:ckraymond@u.washington.edu)  
 Class: shotgun.

## COMMENT

This GSS is part of a random genomic sequencing program of thirteen

## BASE COUNT

242 a 319 c 344 g 199 t 7 others

ORIGIN

Query Match 7.2% Score 19; DB 29; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 3e+22;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CGTGTGCTGTAAGATCACC 144  
 |||||  
 Db 245 CGTGTGCTGTAAGATCACC 227

RESULT 12  
 AA480437 102 bp mRNA linear EST 15-AUG-1997  
 LOCUS ne0a12.s: NCI\_CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:909598, mRNA  
 DEFINITION sequence.  
 ACCESSION AA480437 GI:2208588  
 VERSION AA480437  
 KEYWORDS EST.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 : (bases 1 to 102)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-rtm1@nih.gov](mailto:cgabs-rtm1@nih.gov)  
 Tissue Procurement: Lee Helmar, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the T.M.A.C.E. Consortium/UM/LLNL at:  
[www.bio.livn.gov/bdip/image/image.html](http://www.bio.livn.gov/bdip/image/image.html)  
 Seq primer: 41m13 fwd. ET from Amerstam  
 High quality sequence stop: 77.  
 Location/Qualifiers  
 1..102  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:909598"  
 /issue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Ew1"  
 /note="Vector: pAMP10; mRNA made from Ewing's sarcoma.  
 cDNA made by oligo-dT priming. Non-directionally cloned.  
 Size selected on agarose gel. Average insert size 600 bp.  
 Reference: Krizman et al. (1996) Cancer Research  
 56:5380-5383."

BASE COUNT 23 a 30 c 26 g 23 t

ORIGIN

Query Match 6.8% Score 18; DB 9; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGCTGTAAGATCACC 144  
 |||||  
 Db 53 GTGCTGCTGTAAGATCACC 36

RESULT 13  
 AA706470 140 bp mRNA linear EST 24-DEC-1997  
 LOCUS ag95c07.s1 StrataGene hnt neuron (#937233) Homo sapiens cDNA clone  
 DEFINITION IMAGE:1142220 3', mRNA sequence.  
 ACCESSION AA706470  
 VERSION AA706470 GI:2716388

KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 : (bases 1 to 140)  
 H-11et.D., Allen.M., Rowles.T., Dubouche.T., Gesel.J., Post.S.,  
 Krizman.D., Kucaba.T., Lacy.M., Lennon.G., Marra.M., Martin  
 J., Moore.B., Scheinberg.K., Seppoe.V., Tan.F., Treising.B.,  
 White.Y., Wylie.T., Watserson.R. and Wilson.R.  
 was0-NCI human EST Project  
 Unpublished  
 Contact: Wilson R.  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@wustl.edu](mailto:est@wustl.edu)  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: 40m13 fwd. ET from Amerstam  
 High quality sequence stop: 134.  
 Location/Qualifiers  
 1..140  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1142220"  
 /dev\_stage="hnt neurons"  
 /lab\_host="SCL (Kanamycin resistant)"  
 /clone\_lib="StrataGene hnt neuron (#937233)"  
 /note="Vector: pRiscrypt SK-; Site 1: EcoRI; Site 2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT.  
 Differentiated, post mitotic hnt neurons. Average insert  
 size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
 CATTTCGACACGAG 3' 3' adaptor sequence: 5'  
 CTCAGTTTCTTTCTTTTCTTT 3'

BASE COUNT 32 a 41 c 33 g 34 t

ORIGIN

Query Match 6.8% Score 18; DB 9; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGCTGTAAGATCACC 144  
 |||||  
 Db 53 GTGCTGCTGTAAGATCACC 36

RESULT 14  
 AV940990 141 bp mRNA linear EST 18-JAN-2002  
 LOCUS AV940990 X. Sato unpublished cDNA library, strain H602 adult,  
 DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
 cDNA clone bat10k10 5', mRNA sequence.  
 ACCESSION AV940990 GI:18236787  
 VERSION AV940990.1  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. spontaneum  
 ORGANISM Hordeum vulgare subsp. spontaneum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;  
 Triticeae; Hordeum.  
 : (bases 1 to 141)  
 / Triticeae; Hordeum.  
 Reference Sato,K., Saito,D. and Takega,K.  
 Authors Bailey EST sequencing project in NIG and Okayama Univ  
 Title Unpublished  
 Journal Contact: Tadashi Shin-i  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6656  
 Fax: 81-559-81-6855

FEATURES Email: tshin@genes.nig.ac.jp.  
Location: Qualifiers

1.141

/organism="Hordium vulgare subsp. spontaneum"

/mol\_type="mRNA"

/strain="H602"

/db\_xref="taxon:77009"

/clone="hah30x10"

/tissue\_type="top three leaves"

/dev\_stage="adult, heading stage"

/clone\_lib="X. Sato unpublished cDNA library, strain H602"

adult, heading stage top three leaves"

31 c 44 g 18 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 68 AGCAGCTTCAGGTCGCGC 85

196 AGCAGCTTCAGGTCGCGC 213

192 bp mRNA linear EST 20-MAY-2002

RC4-HT111-22-200-022-ec7 HT111 Homo sapiens cDNA, mRNA sequence.

RC4-54983

RC4-54983

RC4-54983.1 GI:21029039

EST

EST

Homo sapiens (human)

Homo sapiens

Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eumetazoa; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 192)

Dias Neto, E., Garcia, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Strunstein, A., de Oliveira, P.S., Bucher, P., Cornejo, C.V., O'Hare

, M.C., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.C.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
http://www.ludwig.org.br/scripts/getcdna.pl?lib=RC4&c2=RC4-HT111-  
221200-022-e07&c3=2000-12-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 60  
High quality sequence stop: 96.  
Location/Qualifiers  
1.192  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT111"  
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI. A mini-library was made by cloning products  
derived from ORESTES PCR (O.S. Lenters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions.

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 5.8e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 95 TGTGCTACGTGAGACTTGG 112

95 TGTGCTACGTGAGACTTGG 112

35 TGTGCTACGTGAGACTTGG 52

Search completed: November 7, 2003, 14:47:14

Job time: 1320 secs



PI Fadiat M, Beachy KN, Fauquet CM;  
 XX WP1: 1999-551037/46.  
 DR P-PSDB: AAY31969.  
 XX  
 PT Producing plants resistant to single stranded DNA virus, using a single  
 ET stranded DNA-binding protein of the Inoviridae virus family  
 XX  
 PS Claim 9; Page 61; 69pp; English.

The invention relates to a method for producing in a plant geminivirus  
 CC resistance to a single stranded (ss) DNA virus. The method comprises  
 CC introducing a gene capable of expressing a ssDNA-binding protein of the  
 CC inoviridae virus (IVV) family into the plant. The IVV family virus is  
 CC selected from noviviruses or Plectrovirus generaes where the inovirus genus  
 CC virus is selected from a group consisting of Coliphage, enterobacteria  
 CC phage, Pseudomonas phage, Vibrio phage or Xanthomonas phage species. The  
 CC ssDNA-binding protein is a Coliphage coat protein or a gene 5 protein  
 CC and especially a Coliphage M13 gene 5 protein. The ssDNA-binding protein  
 CC of the IVV family interferes with virus spread during the infection  
 CC process of plant viruses of the ssDNA type. By inhibiting virus spread,  
 CC the virus infection is reduced and/or blocked, thereby increasing plant  
 CC resistance to the virus infection. The method can be used for protecting  
 CC plants against e.g. Bayra streak virus, Bean yellow dwarf virus, Bromus  
 CC striate mosaic virus, Chickpea chlorotic dwarf virus, Chloris striate  
 CC mosaic virus, Digitaria streak virus, Digitaria striate mosaic virus,  
 CC Wheat streak virus, Miscanthus streak virus, Tobacco yellow dwarf virus,  
 CC Wheat dwarf virus, Beet curly top virus, Horseadash curly top virus,  
 CC Tomato leafroll virus, Tomato pseudo-curly top virus, Abutilon mosaic  
 CC virus, Acalypha yellow mosaic virus, cassava mosaic virus, Tomato leaf  
 CC curl virus, Cotton leaf curl virus, Banana bunchy top virus, Coconut  
 CC foliar decay virus, Papayan necrotic yellows virus, Milk vetch dwarf  
 CC virus or Subterranean an clover stunt virus. The present sequence  
 CC represents a synthetic DNA sequence of Coliphage M13 gene 5 protein.

Sequence 264 BP; 54 A; 93 C; 86 G; 37 T; 0 other;

Query Match 100.0%; Score 264; DB 20; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 4, 4e-111;  
 Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGGATCAACCCAGGCCGCGTTCACCCGCGAGCGCGTGAAC 60  
 Dc 1 AGATCAAGGTGGATCAACCCAGGCCGCGTTCACCCGCGAGCGCGTGAAC 60  
 QY 61 CGCCAGGCGCAAGCCCTACACGCTGAGAGAGCTGTCTACTGTGACCTGCGCAAGAG 120  
 Dc 61 CGCCAGGCGCAAGCCCTACAGCTGAGAGAGCTGTCTACTGTGACCTGCGCAAGAG 120  
 QY 121 TACCCCTGTGCTGTGTAAGATCACTCTGAGCGAGCGCCGCTTACGCCCGGCTG 180  
 Dc 121 TACCCCTGTGCTGTGTAAGATCACTCTGAGCGAGCGCCGCTTACGCCCGGCTG 180  
 QY 181 TACACCGGTGACCTGAGAGAGCTTCAAGGTGGCCAG-TTGGCAGCTCGATGAGACGC 240  
 Dc 181 TACACCGGTGACCTGAGAGAGCTTCAAGGTGGCCAG-TTGGCAGCTCGATGAGACGC 240  
 QY 241 CTGGCCCTGTGCTGCGCCCAAGTAA 264  
 Dc 241 CTGGCCCTGTGCTGCGCCCAAGTAA 264

RESULT 2  
 ABL15453  
 ID ABL15453 standard; cDNA; 3596 BP.  
 XX  
 AC ABL15453;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40941.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.  
 XX  
 CS Drosophila melanogaster.  
 XX  
 XX WO200171042-A2.  
 XX  
 XX 27-SEP-2001.

XX 23-MAR-2002; 2001WO-US09231.  
 XX  
 XX 23-MAR-2002; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 DR P-PSDB: ABB71350.  
 XX  
 XX (EPKE ) PE CORP NY.  
 PA  
 XX  
 XX Venter CC, Adams M, Li FWD, Myers EW;  
 PI  
 XX WP1: 2001-656869/75.  
 DR P-PSDB: ABB71350.

PI New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 40941; 21pp - Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutic and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB130840-AB16175) and the encoded proteins  
 CC (AB13717-AB172072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WFO  
 CC at ftp://wfo.int/pub/pub/seqs\_per\_sequences.

Sequence 3596 BP; 961 A; 912 C; 886 G; 911 T; 0 other;

Query Match 7.24%; Score 19; DB 23; Length 3596;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 GATCAACCTGTGACGAGGCG 196  
 Dc 1453 GATCAACCTGTGACGAGGCG 1451

RESULT 3  
 AB12454/C  
 ID AB12454 standard; cDNA; 5892 BP.  
 XX  
 AC AB12454;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31844.  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX  
 KW pharmaceutical; gene; ss.  
 XX  
 XX Drosophila melanogaster.  
 CS  
 XX  
 XX WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 XX  
 XX 23-MAR-2002; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX



PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249224.  
 PR 17-NOV-2000; 2000US-0249225.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250163.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 03-DEC-2000; 2000US-0251030.  
 PR 03-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 06-DEC-2000; 2000US-0251658.  
 PR 06-DEC-2000; 2000US-0251659.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254397.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 FA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 FT Rosen CA, Barash SC, Ruben SX;  
 XX  
 DR WPI: 2001-465576/50.  
 XX  
 FT Isolated nucleic acid molecule encoding a reproductive system antigen.  
 FT is used in preventing, treating or ameliorating a medical condition.  
 XX  
 PS Disclosure: SEQ ID NO 7092; 1297bp - Sequence Listing: English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a genomic sequence encoding a  
 CC protein of the invention.  
 XX  
 SQ Sequence 14769 BP; 3721 A; 3639 C; 3593 G; 3816 T; 0 other;

Query Match 7.28; Score 19; DB 22; Length 14769;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123' CCCCCTGCTGCTGAGATC 141  
 |||||||

DB 3773 CCCCCTGCTGCTGAGATC 8791  
 RESULT 5  
 ID ABL15452 standard; CDNA; 26542 BP.  
 XX  
 AC ABL15452;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40818.  
 XX  
 KW Drosophila; developmental biology; cell signaling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WC020171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001MO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-1916372.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE) PB CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR P-PSDB; ABB71349.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Claim 1: SEQ ID NO 40818; 21bp - Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL15452-AB15453) expressed DNA  
 CC sequences (AB15454-AB15455) and the encoded proteins.  
 CC (AB15457-AB15458).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 CC  
 SQ Sequence 26542 BP; 8259 A; 5420 C; 5350 G; 7513 T; 0 other;

Query Match 7.28; Score 19; DB 22; Length 26542;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 GATCACCCTGAGAGAGG 156  
 |||||||

RESULT 6  
 ID AAC37969  
 AC AAC37969;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 19319.



KM Hybridisation assay; Genetic mapping; Gene expression control;  
KM Protein identification; Signal transduction pathway;  
KM Metabolic pathway; Promoter; Termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 200CEP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0122548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 29-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130549.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134376.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138543.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142190.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143524.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145312.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145319.  
PR 28-JUL-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0145386.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147392.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147307.  
PR 06-AUG-1999; 99US-0147416.  
PR 08-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 22-AUG-1999; 99US-0149902.  
PR 22-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151103.  
PR 31-AUG-1999; 99US-0151103.

PR 01-SEP-1999; 99US-015193C.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153759.  
PR 15-SEP-1999; 99US-0154015.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154719.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156598.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 28-OCT-1999; 99US-0161994.  
PR 29-OCT-1999; 99US-0162142.

Query Match: 6.8%; Score: 18; DB: 21; Length: 487;  
Best Local Similarity: 100.0%; Pred. No. 1e-02; 0; Indels: 0; Gaps: 0;

Matches: 18; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

CY 177 CCGTACACCGCTGACCT 194  
          |||||  
Db 93 CCGTACACCGCTGACCT 110

RESULT 7

AAC46777

ID AAC46777 standard; DNA, 586 BP.

AC AAC46777;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51363.

KM Hybridisation assay; genetic mapping; gene expression control;

KW Protein identification; signal transduction pathway;

XX Metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.  
PR 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
XX  
PR 25-FEB-1999; 99US-0121525.  
PR 05-MAR-1999; 99US-0121580.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128274.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132046.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132563.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134259.  
PR 14-MAY-1999; 99US-0134262.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134376.  
PR 13-MAY-1999; 99US-0134394.  
PR 21-MAY-1999; 99US-0135194.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137628.  
PR 04-JUN-1999; 99US-0137802.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138240.  
PR 13-JUN-1999; 99US-0138567.  
PR 14-JUN-1999; 99US-0139219.  
PR 16-JUN-1999; 99US-0139432.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139482.  
PR 18-JUN-1999; 99US-0139484.  
PR 18-JUN-1999; 99US-0139485.  
PR 18-JUN-1999; 99US-0139486.  
PR 18-JUN-1999; 99US-0139487.  
PR 18-JUN-1999; 99US-0139488.  
PR 19-JUN-1999; 99US-0139489.  
PR 22-JUN-1999; 99US-0140303.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140635.  
PR 26-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.

```

PR 02-JUL-1999: 99US-0143055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144684.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 28-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147322.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 10-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149829.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150565.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 30-AUG-1999: 99US-0151338.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 13-SEP-1999: 99US-0154718.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.

```

```

PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156556.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157866.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158169.
PR 13-OCT-1999: 99US-0158293.
PR 13-OCT-1999: 99US-0158294.
PR 13-OCT-1999: 99US-0158295.
PR 14-OCT-1999: 99US-0158296.
PR 14-OCT-1999: 99US-0158329.
PR 14-OCT-1999: 99US-0158330.
PR 14-OCT-1999: 99US-0158331.
PR 14-OCT-1999: 99US-0158637.
PR 14-OCT-1999: 99US-0158638.
PR 18-OCT-1999: 99US-0158584.
PR 21-OCT-1999: 99US-0160724.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 22-OCT-1999: 99US-0160770.
PR 22-OCT-1999: 99US-0160814.
PR 22-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160983.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 26-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161921.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

```

Query Match Best Local Similarity 6.6% Score 18: FB 21: Length 586:  
Matches 18: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

Cy 177 CCTGTACACCGTCACT 154
Db 149 CCTGTACACCGTCACT 156

```

```

RESULT 8
AAH13510/c
ID AAH13510 standard cDNA, 587 BP.
XX
AC AAH13510;
XX
DT 26-SEP-2001 (first entry)
DE Human cDNA clone (3'-primer) SEQ ID NO:10365.
XX
KM Human (primer) detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX
XX EPI07467-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999: 99JP-0248036.
PR 27-AUG-1999: 99JP-0300253.
PR 11-JAN-2000: 2000JP-0118776.
PR 02-MAY-2000: 2000JP-0183767.

```

```

PR 09-JUN-2000; 2000CP-024:899.
XX (HELI-) HELIX RES INST.
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-3-8749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 3: SEQ ID 10365; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH3166 to AAH3628 and
XX AAH1633 to AAH18742 represent human cDNA sequences; AAB9446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 587 BP; 137 A; 152 C; 126 G; 165 T; 7 other;
SQ
XX
XX Query Match 6.8%; Score 18; DB 2; Length 587;
XX Best Local Similarity 100.0%; Pred. No. 1e-02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 127 GTGCTGCTGAGATCACC 144
XX |||||
XX 295 GTGCTGCTGAGATCACC 276
XX
XX RESULT 9
XX ABQ66303/c
XX ID ABQ66303 standard; DNA; 610 BP.
XX
XX AC ABQ66303.
XX
XX 21-AUG-2002 (first entry);
XX
XX Arabidopsis thaliana polynucleotide SEQ ID NO 880.
XX
XX Arabidopsis thaliana; thale cress; plant; transgenic; GYO; disease;
XX stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;
XX insecticide; antibiotic; ds.
XX
XX Arabidopsis thaliana.
XX
XX OS
XX PN US2002059663-A1.
XX
XX 16-MAY-2002.
XX
XX 26-JAN-2002; 2001US-0770149.
XX
XX 27-JAN-2000; 2000US-178506P.

```

```

XX (GRL/) GORLACH J.
XX (ANY/) AN Y.
XX (HAM/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUY/) YU Y.
XX (RAME/) RAMEKA J G.
XX (PAGE/) PAGE A.
XX (MATH/) MATHW A V.
XX (LEDF/) LEDFORD B L.
XX (WOES/) WOESSNER J P.
XX (HAAS/) HAAS W D.
XX (GARC/) GARCIA C A.
XX (KRIC/) KRICKER M.
XX (SLAT/) SLATER T.
XX (DAVI/) DAVIS K R.
XX (ALE/) ALLEN K.
XX (HOFF/) HOFFMAN N.
XX (HURB/) HURBAN P.
XX
XX Scrlach J, An Y, Hamilton CM, Price CL, Raines TX, Yu Y,
XX Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,
XX Hurban P;
XX WPI; 2002-479224/51.
XX
XX New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
XX useful e.g. for preparing transgenic plants with increased resistance
XX or altered metabolism -
XX
XX Claim 1; SEQ ID NO 880; 40pp + Sequence Listing; English.
XX
XX The invention relates to nucleic acids (i) that hybridise under stringent
XX conditions to any of 399 sequences (ABQ65424-ABQ66422) or their
XX fragments; (i) are used to express the corresponding polypeptides (ii) or
XX to produce genetically modified plant cells or transgenic plants, which
XX may have improved resistance to disease or stress, or altered
XX metabolic/biosynthetic pathways (for production of commercial
XX nutritional or medicinal products), or generally any trait of interest,
XX or can be used to screen for biologically active agents (e.g. fungicides,
XX insecticides and antibiotics).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO at seqdata.uspto.gov/sequence.html?docid=93939770149.
XX
XX Sequence 610 BP; 175 A; 152 C; 151 G; 130 T; 0 other;
SQ
XX
XX Query Match 6.8%; Score 18; DB 2; Length 610;
XX Best Local Similarity 100.0%; Pred. No. 1e-02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 177 CCGTACACCGTCGCACT 194
XX |||||
XX 487 CCGTACACCGTCGCACT 470
XX
XX RESULT 10
XX ACC50489
XX ID ACC50489 standard; cDNA; 1142 BP.
XX
XX AC ACC50489.
XX
XX 12-JUN-2003 (first entry);
XX
XX Human secreted protein coding sequence, SEQ ID 156.
XX
XX Cardiac; antiarrhythmic; antiarteriosclerotic; vasotrophic; cytotstatic;
XX vulnerary; antiinflammatory; neurotrophic; neuroprotective;
XX antiparkinsonian; gene therapy; human; cardiovascular disorder;
XX gene; ss.

```

```

OS Homo sapiens.
XX
XX WO200295010-A2.
XX
XX
XX 28-NOV-2002.
XX
XX 19-MAR-2002: 2002MO-US09785.
XX
XX 21-MAR-2001: 2001US-277340P.
XX 19-JUL-2001: 2001US-306171P.
XX 13-NOV-2001: 2001US-331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2003-129429/12.
XX
XX Novel human secreted proteins, useful for detecting, preventing,
XX diagnosing, prognosticating, treating and/or ameliorating
XX cardiovascular disorders such as arrhythmia -
XX
XX Claim 21: SEQ ID 156: 1881bp; English.
XX
XX The present invention relates to novel human secreted proteins
XX (ABR47693-ABR48145) and their coding sequences (ACC50344-ACC50855). The
XX proteins and their coding sequences are useful for the preparation of a
XX diagnostic or pharmaceutical composition for diagnosing or treating a
XX cardiovascular disorder (e.g., arrhythmia, tachycardia, cardiac arrest,
XX coronary arteriosclerosis and myocardial ischemia), neural disorders,
XX immune system disorders, muscular disorders, reproductive disorders,
XX gastrointestinal disorders, pulmonary disorders, renal disorders,
XX proliferative disorders and/or cancerous diseases and conditions, for
XX wound healing and epithelial cell proliferation, to treat inflammation or
XX infection, for treating thrombosis and arteriosclerosis, for treating or
XX preventing neural damage which occurs in neuronal disorders or
XX neurodegenerative conditions such as Alzheimer's disease and Parkinson's
XX disease, to enhance bone and periodontal regeneration and aid in tissue
XX transplants or bone grafts, to prevent skin aging or hair loss, to
XX stimulate growth and differentiation of hematopoietic cells and bone
XX marrow cells when used in combination with other cytokines, to maintain
XX organs before transplantation or for supporting cell culture of primary
XX tissues, to increase or decrease differentiation or proliferation of
XX embryonic stem cells, or to modulate mammalian characteristics or
XX metabolism.
XX Note: The sequence data for this patent was published in electronic
XX format and is available from WIPO at
XX ftp://wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1142 BP: 335 A; 246 C; 295 G; 266 T; 0 other:
XX
XX Query Match: 6.8%; Score 19; DB 25; Length 1142;
XX Best Local Similarity 100.0%; Pred. No. 98;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 127 GTGCTGCTGAGATCACC 144
XX 790 GTGCTGCTGAGATCACC 807
XX
XX RESULT 11
XX ID ABR71271 standard; cDNA: 1142 BP.
XX
XX AC ABR71271;
XX
XX 03-APR-2003 (first entry)
XX
XX Human secreted protein-encoding gene #2 cDNA clone HB31557, SEQ ID NO:92.
XX
XX DE Human secreted protein-encoding gene #2 cDNA clone HB31557, SEQ ID NO:92.
XX
XX KM Human: secreted protein; digestive disorder; gastro-intestinal disorder;
XX mouth; oesophagus; stomach; small intestine; large intestine; liver;
XX biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;
XX

```

```

XX immune disorder; inflammation; infection; wound healing; drug screening;
XX chromosome identification; chromosome mapping; cytostatic;
XX anti-inflammatory; immunosuppressive; antiviral; gene therapy; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200276488-A1.
XX
XX
XX 03-OCT-2002.
XX
XX 19-MAR-2002: 2002MO-US08276.
XX
XX 21-MAR-2001: 2001US-277340P.
XX 19-JUL-2001: 2001US-306171P.
XX 13-NOV-2001: 2001US-331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2003-029900/02.
XX
XX P-PSDB: ABR00092.
XX
XX New human secreted proteins and nucleic acids, useful for detecting,
XX preventing, diagnosing, prognosticating, treating and/or ameliorating
XX e.g. gastrointestinal diseases and disorders, or cancers -
XX
XX Claim 21: Page 807-808: 1216bp; English.
XX
XX ABR71130-ABR71478 represent cDNA corresponding to the human secreted
XX protein genes, and ABR0011-ABR0029 represent the proteins they encode.
XX ABR71479-ABR71540 represent human secreted protein coding fragments. The
XX invention also encompasses antibodies specific for the secreted proteins,
XX the use of the secreted proteins in drug screening, and treatment. The
XX vectors and host cells comprising a nucleic acid of the invention. The
XX secreted proteins, nucleic acids encoding them, antibodies or antibody
XX fragments specific for the secreted proteins, and modulators of protein
XX activity are useful for diagnosing, treating, ameliorating or preventing
XX digestive disorders. Such conditions include disorders of the mouth,
XX oesophagus, stomach, small intestine, large intestine, liver, biliary
XX tract and pancreas, and include cancers of these organs and tissues. The
XX secreted proteins and their nucleic acids may also be used in the
XX treatment of immune disorders, inflammation, infection,
XX hyperproliferative disorders, and to promote wound healing. Nucleic acids
XX of the invention may be used for chromosome identification, chromosome
XX mapping, in gene therapy, for identifying individuals from minute
XX biological samples, as hybridisation probes, and as molecular weight
XX markers. The present sequence represents a human secreted protein-
XX encoding cDNA clone of the invention.
XX
XX Sequence 1142 BP: 335 A; 246 C; 295 G; 266 T; 0 other:
XX
XX Query Match: 6.8%; Score 19; DB 25; Length 1142;
XX Best Local Similarity 100.0%; Pred. No. 98;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 127 GTGCTGCTGAGATCACC 144
XX 790 GTGCTGCTGAGATCACC 807
XX
XX RESULT 12
XX ID ABR35301 standard; cDNA: 1265 BP.
XX
XX AC ABR35301;
XX
XX 08-MAY-2002 (first entry)
XX
XX Human cDNA encoding secreted protein #439.
XX
XX DE Human cDNA encoding secreted protein #439.
XX
XX KM Human: secreted protein; gene; ss; nutritional supplement; haemophilia;
XX viral infection; bacterial infection; fungal infection; diabetes; asthma;
XX

```

KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;  
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;  
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;  
 KW tissue regeneration; wound healing; burn; haematopoiesis;  
 KW myeloid cell deficiency; lymphoid cell deficiency.  
 XX  
 OS Homo sapiens.  
 PN WO200177288-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 29-MAR-2001; 2001WO-US10224.  
 XX  
 PR 06-APR-2000; 2000US-195582P.  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ,  
 PI Gullotta K, Graham JR;  
 DR WPI; 2002-179321/23.  
 XX  
 PT Five hundred and ninety two polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders  
 PT  
 PS Claim 1; Page 299; 372pp; English.  
 XX  
 CC The invention relates to 592 polynucleotides which have been derived from  
 CC a variety of human tissue sources and which encode novel secreted  
 CC proteins. The polynucleotides can be used as probes for the  
 CC identification and isolation of full length cDNA and genomic DNA. The  
 CC polynucleotides and proteins can also be used as nutritional supplements.  
 CC The proteins are useful in the treatment of various immune deficiencies.  
 CC and disorders such as viral infections, bacterial infections, fungal  
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple  
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions  
 CC and conditions (e.g. asthma). They are also useful for treating  
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
 CC useful for tissue regeneration, for wound healing and in the treatment of  
 CC burns, incisions and ulcers. The proteins are also useful for regulating  
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.  
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.  
 CC  
 XX  
 SQ Sequence 1265 BP; 368 A; 278 C; 308 G; 311 T; 0 other;  
 Query Match 6.8%; Score 18; DB 24; Length 1265;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 127 GTGCTGTGAAGATCACC 144  
 |||||  
 Db 971 GTGCTGTGAAGATCACC 986  
 |||||  
 RESULT 13  
 AAX99154  
 ID AAX99154 standard; DNA; 1780 BP.  
 XX  
 AC AAX99154;  
 XX  
 DT 24-SEP-1999 (first entry);  
 XX  
 DE DNA encoding N. meningitidis protein ORF116.  
 XX  
 KW Neisseria meningitidis protein; pharmaceutical; vaccine; diagnosis;  
 KW bacterial infection; treatment; ss.  
 XX

OS Neisseria meningitidis.  
 XX  
 PN WO9936344-A2.  
 XX  
 PD 22-JUL-1999.  
 XX  
 PF 14-JAN-1999; 99WO-IB00103.  
 XX  
 PR 09-OCT-1998; 98GB-0022143.  
 XX  
 PR 14-JAN-1998; 98GB-0000760.  
 XX  
 PR 01-SEP-1998; 98GB-0019015.  
 XX  
 PA (CHIRON) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizsa X, Rappuoli R, Scarlato V;  
 XX  
 DR WPI; 1999-444400/37.  
 XX  
 DR P-2529; AAY27232.  
 XX  
 PT New protein and its nucleotide sequence, useful in vaccines or  
 PT diagnostic compositions for treating and/or preventing Neisseria  
 PT meningitidis infections  
 PT  
 PS Claim 9; Page 95; 123pp; English.  
 XX  
 CC The invention provides proteins AAY27232-245; from Neisseria  
 CC meningitidis (strains A and B) and nucleic acid sequences (AAX99123-167)  
 CC encoding the proteins. Compositions comprising the protein, nucleic acid  
 CC or antibody specific to the protein are useful as pharmaceuticals, e.g. a  
 CC vaccine composition or a diagnostic composition. The composition is also  
 CC useful for treating or preventing an infection due to Neisseria  
 CC bacteria, especially Neisseria meningitidis.  
 CC  
 XX  
 SQ Sequence 1760 BP; 577 A; 477 C; 431 G; 285 T; 10 other;  
 Query Match 6.8%; Score 19; DB 20; Length 1760;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 CACCACCGGCGGCGGT 56  
 |||||  
 Db 1615 CACCACCGGCGGCGGT 1632  
 |||||  
 RESULT 14  
 ID AAX81446 standard; DNA; 1782 BP.  
 XX  
 AC AAX81446;  
 XX  
 DT 04-DEC-2000 (first entry);  
 XX  
 DE N. meningitidis Memb polynucleotide sequence GRF number 117.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; Mening. ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0133794.  
 XX  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIRON) CHIRON CORP.  
 XX  
 PI Frazer CW, Hickey E, Peterson C, Tettelin H, Wenter CT,  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarlato V;  
 XX

PI Rappaport R. Pizsa M.  
 XX  
 DR WPI: 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea*.  
 XX  
 PS Disclosure, Page 229, 1760pp: English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81451 to AAA82414  
 CC represent specifically obtained *Neisseria meningitidis* genomic DNA  
 CC sequences. AAA81260 to AAA81333 and AAA82562 to AAA82563 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins. AAA82254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament for in the manufacture of a  
 CC medicament for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus* B; against all serotypes;  
 CC and/or against all pathogenic *Neisseria*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus* B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 SQ Sequence 1782 BP; 577 A; 477 C; 431 G; 285 T; 12 other;  
 XX  
 QY Query Match 6.8%; Score 18; DB 21; Length 1782;  
 Db Best Local Similarity 100.0%; Pred. NO. 96;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 CACCACCCGAGCGGCGT 56  
 Db 1617 CACCACCCGAGCGGCGT 1634  
 XX  
 RESULT 15  
 AAH14055  
 ID AAH14055 standard; cDNA: 1927 BP.  
 XX  
 AC AAH14055;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:1195.  
 XX  
 KW Human; primer; detection; diagnosis; anti-sense therapy; gene therapy; ss.  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2003; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 22-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX

PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto G;  
 PI Ishii S, Sugiyama T, Wakatsuki A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318745/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs.  
 XX  
 PS Claim 8, SEQ ID 1195; 2537bp - CD ROM: English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5622 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH13166 to AAH13168 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH93893 represent human amino acid sequences; and AAH13623 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1927 BP; 484 A; 453 C; 538 G; 452 T; 0 other;  
 XX  
 QY Query Match 6.8%; Score 18; DB 22; Length 1927;  
 Db Best Local Similarity 100.0%; Pred. NO. 95;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 127 GTGCTGTGAGATGATCACC 144  
 Db 1633 GTGCTGTGAGATGATCACC 1650  
 XX

Search completed: November 7, 2003, 13:55:54  
 Job time : 195 secs

[illegible]



Db 48 CCCCCTGCTGTGAGATC 30

# RESULT 2

US-10-027-632-157148/c

Sequence 157148, Application US/10027632

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,576

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 157148

LENGTH: 815

TYPE: DNA

ORGANISM: Human

US-10-027-632-157148

Query Match 7.2%; Score 19; DB 11; Length 815;

Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCCCCTGCTGTGAGATC 141

Db 48 CCCCCTGCTGTGAGATC 30

## RESULT 3

US-10-156-761-2399/c

Sequence 2399, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 2399

LENGTH: 849

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(849)

US-10-156-761-2399

Query Match 7.2%; Score 19; DB 14; Length 649;

Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 GGTGGGCACTTCGACG 225

Db 664 GGTGGGCACTTCGACG 646

## RESULT 4

US-09-764-891-7092

Sequence 7092, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antipodes

FILE REFERENCE: PCC06

CURRENT APPLICATION NUMBER: US/09/764,891

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: Patentn Ver. 2.0

SEQ ID NO 7092

LENGTH: 14769

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-891-7092

Query Match 7.2%; Score 19; DB 11; Length 14769;

Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CCCCCTGCTGTGAGATC 141

Db 8773 CCCCCTGCTGTGAGATC 8791

## RESULT 5

US-10-156-761-1

Sequence 1, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 1

LENGTH: 925608

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS feature

LOCATION: (1487715)

OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Query Match 7.2%; Score 19; DB 14; Length 9025608;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 GGTGGCCAGTTCGCAGC 225  
|||||  
Db 2939921 GGTGGCCAGTTCGCAGC 2939939

## RESULT 6

US-09-770-149-880/C  
Sequence 880, Application US/09770149  
Patent No. US20020059663A1  
GENERAL INFORMATION:  
APPLICANT: Goriach, Jörn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Kameoka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kicker, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hubbar, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2024 (PAPA-013PPV)  
CURRENT APPLICATION NUMBER: US/09770149  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/1178,506  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 880  
LENGTH: 610  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-149-880

Query Match: 6.8%; Score 18; DB 9; Length 610;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 177 CCTGTACCGTGCACCT 194  
|||||  
Db 487 CCTGTACCGTGCACCT 470

## RESULT 7

US-09-822-849A-439  
Sequence 439, Application US/09822849A  
Patent No. US20020045170A1  
GENERAL INFORMATION:  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Agostino, Michael C.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulukota, Kamaakar  
APPLICANT: Graham, James R.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6403  
CURRENT APPLICATION NUMBER: US/09/822,849A  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: 60/195,582  
PRIOR FILING DATE: 2000-04-06

NUMBER OF SEQ ID NOS: 598  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 439  
LENGTH: 1265  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-849A-439

Query Match: 6.8%; Score 18; DB 9; Length 1265;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAGATCACC 144  
|||||  
Db 971 GTGCTGTGAGATCACC 938

## RESULT 8

US-10-156-761-5596/C  
Sequence 5596, Application US/10156761  
Publication No. US20030119016A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO: 5596  
LENGTH: 1437  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11...11437)  
US-10-156-761-5596

Query Match: 6.8%; Score 18; DB 14; Length 1437;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GACGAGGCCAGCCGCC 165  
|||||  
Db 159 GACGAGGCCAGCCGCC 142

## RESULT 9

US-10-198-846-10379  
Sequence 10379, Application US/10198846  
Publication No. US20030099744A  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
FILE REFERENCE: WRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18

```
? NUMBER OF SEQ ID NOS: 14084
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 10379
? LENGTH: 2522
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: 1_2_3_4_5_6_7_8_9_10_11_12_13_2519_2520_2521
? OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10379
```

```
Query Match          6.8% Score 18; DB 14; Length 2522;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      127 GTGCTGGTGAAGATCACC 144
DB      2175 GTGCTGGTGAAGATCACC 2192
```

```
RESULT 10
US-09-712-363-31
? Sequence 31, Application US/09712363
? Patent No. US20020164588A1
? GENERAL INFORMATION:
? APPLICANT: Eisenberg, David
? APPLICANT: Rotstein, Sergio H.
? APPLICANT: Marcotte, Edward M.
? TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
? FILE REFERENCE: 07419-032001
? CURRENT APPLICATION NUMBER: US/09/712.363
? PRIOR FILING DATE: 2000-11-13
? PRIOR APPLICATION NUMBER: PCT/US00/02246
? PRIOR FILING DATE: 2000-01-28
? PRIOR APPLICATION NUMBER: 60/179,531
? PRIOR FILING DATE: 2000-02-01
? PRIOR APPLICATION NUMBER: 60/117,844
? PRIOR FILING DATE: 1999-01-29
? PRIOR APPLICATION NUMBER: 60/118,236
? PRIOR FILING DATE: 1999-02-01
? PRIOR APPLICATION NUMBER: 60/126,593
? PRIOR FILING DATE: 1999-03-26
? PRIOR APPLICATION NUMBER: 60/134,093
? PRIOR FILING DATE: 1999-05-14
? PRIOR APPLICATION NUMBER: 60/134,092
? PRIOR FILING DATE: 1999-05-14
? PRIOR APPLICATION NUMBER: 60/165,124
? PRIOR FILING DATE: 1999-11-12
? PRIOR APPLICATION NUMBER: 60/165,086
? PRIOR FILING DATE: 1999-11-12
? NUMBER OF SEQ ID NOS: 292
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 31
? LENGTH: 3951
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
US-09-712-363-31
```

```
Query Match          6.8% Score 18; DB 14; Length 3951;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      41 CCACCCGACGCGGTGA 58
DB      2175 CCACCCGACGCGGTGA 2185
```

```
RESULT 11
US-10-234-007-28/c
? Sequence 28, Application US/10234007
```

```
? Publication No. US2003032160A1
? GENERAL INFORMATION:
? APPLICANT: Hattaboer, Guus
? APPLICANT: Verhulst, Karine Cornelia
? APPLICANT: Schouten, Gerrit Johan
? APPLICANT: Uytendaele, Alphonsus Gerardus Cornelis Maria
? APPLICANT: Bout, Abraham
? TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN A HUMAN CELL
? FILE REFERENCE: 4038,1US
? CURRENT APPLICATION NUMBER: US/10/234,007
? CURRENT FILING DATE: 2002-09-03
? PRIOR APPLICATION NUMBER: US/09/549,463
? PRIOR FILING DATE: 2000-04-14
? PRIOR APPLICATION NUMBER: 06/129,452
? PRIOR FILING DATE: 1999-04-15
? NUMBER OF SEQ ID NOS: 32
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO: 28
? LENGTH: 29
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: oligonucleotide, synthesized sequence, PCR product generate
US-10-234-007-28
```

```
Query Match          6.4% Score 17; DB 14; Length 29;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 CAAGGTGAGATCAAGC 22
DB      26 CAAGGTGAGATCAAGC 10
```

```
RESULT 12
US-10-010-942B-20/c
? Sequence 20, Application US/10010942B
? Publication No. US20030165496A1
? GENERAL INFORMATION:
? APPLICANT: Basu, Gurij
? APPLICANT: Vednack, Ted
? TITLE OF INVENTION: HUMANIZED ANTIBODIES THAT RECOGNIZE
? FILE REFERENCE: ELN-002
? CURRENT APPLICATION NUMBER: US/10/010,942B
? CURRENT FILING DATE: 2002-12-06
? PRIOR APPLICATION NUMBER: US 60/251,892
? PRIOR FILING DATE: 2003-12-06
? NUMBER OF SEQ ID NOS: 63
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO: 20
? LENGTH: 142
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: primer
US-10-010-942B-20
```

```
Query Match          6.4% Score 17; DB 14; Length 142;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 CAAGGTGAGATCAAGC 22
DB      36 CAAGGTGAGATCAAGC 20
```

```
RESULT 13
US-10-198-846-13944
? Sequence 13944, Application US/10198846
? Publication No. US20030099974A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Xille, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Stehrmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-19
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 13944
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-13944

```

```

Query Match      6.4%; Score 17; DB 14; Length 341;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      6 CAAGTGAGATCAAGCC 22
DB      210 CAAGTGAGATCAAGCC 226

```

```

RESULT 14
US-09-918-995-28426
; Sequence 28426, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-32
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 28426
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-28426

```

```

Query Match      6.4%; Score 17; DB 14; Length 458;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      8 AGGTGGAGATCAAGCCC 24
DB      356 AGGTGGAGATCAAGCCC 372

```

```

RESULT 15
US-09-919-580-198
; Sequence 198, Application US/09919560
; Patent No. US2002010832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Changchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121,552
; CURRENT APPLICATION NUMBER: US/09/919,560
; CURRENT FILING DATE: 2001-07-30

```

```

; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 198
; LENGTH: 483
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-198

```

```

Query Match      6.4%; Score 17; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      9 AGGTGGAGATCAAGCCC 24
DB      329 AGGTGGAGATCAAGCCC 345

```

```

Search completed: November 7, 2003, 15:25:46
Job time : 209 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:57:05, Search time 1865 Seconds

(without alignments)  
4694,498 Million Cells updates/sec

Title: US-09-622-500B-3

Perfect score: 264

Sequence: : agatcaaggtgagatcaaa.....gcccggtcccgcccaagtaa 264

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 336368 seqs, 1658189874 residues

Word size: 0

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

```

1: Pending_Patents_NA_Main.*
2: /cgn2_6/ptodata/2/pna/US0986_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US0987_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US0988_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US0989_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US0990_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US0991_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US0992_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US0993_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US0994_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US0995_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US0996_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US0997_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US0998_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US0999_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US1000_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US1001_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US1002_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US1003_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US1004_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US1005_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US1006_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US1007_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US1008_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US1009_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US1010_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US1011_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US1012_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US1013_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US1014_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US1015_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US1016_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US1017_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US1018_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US1019_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US1020_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US1021_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US1022_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US1023_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US1024_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US1025_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US1026_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US1027_COMB.seq.*

```

Result No.	Score	Query Match	Length	DB ID	Description
1	264	100.0	264	27	US-09-622-500B-3
2	20	7.6	1023	53	US-10-437-963-7240C
3	20	7.6	31291	27	US-09-620-392-11855
4	20	7.6	54568	31	US-09-702-134-8795

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

44:	/cgn2_6/ptodata/2/pna/US1028_COMB.seq.*
45:	/cgn2_6/ptodata/2/pna/US1029_COMB.seq.*
46:	/cgn2_6/ptodata/2/pna/US1030_COMB.seq.*
47:	/cgn2_6/ptodata/2/pna/US1031_COMB.seq.*
48:	/cgn2_6/ptodata/2/pna/US1032_COMB.seq.*
49:	/cgn2_6/ptodata/2/pna/US1033_COMB.seq.*
50:	/cgn2_6/ptodata/2/pna/US1034_COMB.seq.*
51:	/cgn2_6/ptodata/2/pna/US1035_COMB.seq.*
52:	/cgn2_6/ptodata/2/pna/US1036_COMB.seq.*
53:	/cgn2_6/ptodata/2/pna/US1037_COMB.seq.*
54:	/cgn2_6/ptodata/2/pna/US1038_COMB.seq.*
55:	/cgn2_6/ptodata/2/pna/US1039_COMB.seq.*
56:	/cgn2_6/ptodata/2/pna/US1040_COMB.seq.*
57:	/cgn2_6/ptodata/2/pna/US1041_COMB.seq.*
58:	/cgn2_6/ptodata/2/pna/US1042_COMB.seq.*
59:	/cgn2_6/ptodata/2/pna/US1043_COMB.seq.*
60:	/cgn2_6/ptodata/2/pna/US1044_COMB.seq.*
61:	/cgn2_6/ptodata/2/pna/US1045_COMB.seq.*
62:	/cgn2_6/ptodata/2/pna/US1046_COMB.seq.*
63:	/cgn2_6/ptodata/2/pna/US1047_COMB.seq.*
64:	/cgn2_6/ptodata/2/pna/US1048_COMB.seq.*
65:	/cgn2_6/ptodata/2/pna/US1049_COMB.seq.*
66:	/cgn2_6/ptodata/2/pna/US1050_COMB.seq.*
67:	/cgn2_6/ptodata/2/pna/US1051_COMB.seq.*
68:	/cgn2_6/ptodata/2/pna/US1052_COMB.seq.*
69:	/cgn2_6/ptodata/2/pna/US1053_COMB.seq.*
70:	/cgn2_6/ptodata/2/pna/US1054_COMB.seq.*
71:	/cgn2_6/ptodata/2/pna/US1055_COMB.seq.*
72:	/cgn2_6/ptodata/2/pna/US1056_COMB.seq.*
73:	/cgn2_6/ptodata/2/pna/US1057_COMB.seq.*
74:	/cgn2_6/ptodata/2/pna/US1058_COMB.seq.*
75:	/cgn2_6/ptodata/2/pna/US1059_COMB.seq.*
76:	/cgn2_6/ptodata/2/pna/US1060_COMB.seq.*
77:	/cgn2_6/ptodata/2/pna/US1061_COMB.seq.*
78:	/cgn2_6/ptodata/2/pna/US1062_COMB.seq.*
79:	/cgn2_6/ptodata/2/pna/US1063_COMB.seq.*
80:	/cgn2_6/ptodata/2/pna/US1064_COMB.seq.*
81:	/cgn2_6/ptodata/2/pna/US1065_COMB.seq.*
82:	/cgn2_6/ptodata/2/pna/US1066_COMB.seq.*
83:	/cgn2_6/ptodata/2/pna/US1067_COMB.seq.*
84:	/cgn2_6/ptodata/2/pna/US1068_COMB.seq.*
85:	/cgn2_6/ptodata/2/pna/US1069_COMB.seq.*
86:	/cgn2_6/ptodata/2/pna/US1070_COMB.seq.*
87:	/cgn2_6/ptodata/2/pna/US1071_COMB.seq.*
88:	/cgn2_6/ptodata/2/pna/US1072_COMB.seq.*
89:	/cgn2_6/ptodata/2/pna/US1073_COMB.seq.*
90:	/cgn2_6/ptodata/2/pna/US1074_COMB.seq.*
91:	/cgn2_6/ptodata/2/pna/US1075_COMB.seq.*
92:	/cgn2_6/ptodata/2/pna/US1076_COMB.seq.*
93:	/cgn2_6/ptodata/2/pna/US1077_COMB.seq.*
94:	/cgn2_6/ptodata/2/pna/US1078_COMB.seq.*
95:	/cgn2_6/ptodata/2/pna/US1079_COMB.seq.*
96:	/cgn2_6/ptodata/2/pna/US1080_COMB.seq.*
97:	/cgn2_6/ptodata/2/pna/US1081_COMB.seq.*
98:	/cgn2_6/ptodata/2/pna/US1082_COMB.seq.*
99:	/cgn2_6/ptodata/2/pna/US1083_COMB.seq.*
100:	/cgn2_6/ptodata/2/pna/US1084_COMB.seq.*
101:	/cgn2_6/ptodata/2/pna/US1085_COMB.seq.*
102:	/cgn2_6/ptodata/2/pna/US1086_COMB.seq.*



```
SEQ ID NO: 11865
LENGTH: 33291
TYPE: DNA
ORGANISM: Oryza sativa
OTHER INFORMATION: unsure at all n locations
US-09-620-392-11865
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 40; Length 33291;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 194 TGAGCAGCTTCAAGTCGCG 213
DB 4348 TGAGCAGCTTCAAGTCGCG 4329
```

```
RESULT 4
US-09-702-134-8795/C
Sequence 8795, Application US/09702134
GENERAL INFORMATION:
APPLICANT: Boukharov, Andrey A.
APPLICANT: Cao, Yongwei
APPLICANT: Kovacic, David K.
APPLICANT: Liu, Jingsong
APPLICANT: McIninch, James
APPLICANT: Wu, Wei
TITLE OF INVENTION: Plant Genome Sequence and Uses Thereof
FILE REFERENCE: 38-21(5:237)F
CURRENT APPLICATION NUMBER: US/09/702,134
CURRENT FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 52202
SEQ ID NO: 8795
LENGTH: 54568
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(54568)
OTHER INFORMATION: unsure at all n locations
US-09-702-134-8795
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 40; Length 54568;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 194 TGAGCAGCTTCAAGTCGCG 213
DB 25625 TGAGCAGCTTCAAGTCGCG 25606
```

```
RESULT 5
US-09-815-264-76662/C
Sequence 76662, Application US/09815264
GENERAL INFORMATION:
APPLICANT: Boukharov, Andrey A.
APPLICANT: Cao, Yongwei
APPLICANT: Dotson, Staton B.
APPLICANT: Koshi, Jeffrey M.
APPLICANT: Kovacic, David K.
APPLICANT: Liu, Jingsong
APPLICANT: McIninch, James
APPLICANT: Wu, Wei
TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
FILE REFERENCE: 38-21(5:1237)G
CURRENT APPLICATION NUMBER: US/09/815,264
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 09/620,392
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: US 09/702,134
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 109669
SEQ ID NO: 76662
LENGTH: 54568
```

```
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(54568)
OTHER INFORMATION: unsure at all n locations
US-09-815-264-76662
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 40; Length 54568;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 194 TGAGCAGCTTCAAGTCGCG 213
DB 25625 TGAGCAGCTTCAAGTCGCG 25606
```

```
RESULT 6
US-60-466-412-371101
Sequence 371101, Application US/60466412
GENERAL INFORMATION:
APPLICANT: CARROLL, Michele
APPLICANT: FAKOBLOVA, Olga
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLOC21466
CURRENT APPLICATION NUMBER: US/60/466,412
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 429241
SOFTWARE: FAST-SEQ for Windows Version 4.0
SEQ ID NO: 371101
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-60-466-412-371101
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1,46+32; Length 201;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 123 CCCCCTGCTGTGAGATC 141
DB 28 CCCCCTGCTGTGAGATC 46
```

```
RESULT 7
US-60-217-080-2972
Sequence 2972, Application US/60217080
GENERAL INFORMATION:
APPLICANT: Green, Matthew
APPLICANT: Stigor, Murray R.
TITLE OF INVENTION: Compositions Isolated from Bovine Tissue
FILE REFERENCE: 105192
CURRENT APPLICATION NUMBER: US/60/217,080
CURRENT FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 35169
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 2972
LENGTH: 215
TYPE: DNA
ORGANISM: Bovine
US-60-217-080-2972
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 1,46+32; Length 215;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 146 TGACGAGGAGCCAGCCGCG 164
DB 94 TGACGAGGAGCCAGCCGCG 102
```

```
RESULT 8
US-60-170-373-934/C
; Sequence 934, Application US/60170373
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
; FILE REFERENCE: C0000149
; CURRENT APPLICATION NUMBER: US/60/170,373
; CURRENT FILING DATE: 1999-12-13
; NUMBER OF SEQ ID NOS: 4282
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 934
; LENGTH: 364
; TYPE: DNA
; ORGANISM: Human
US-60-170-373-934

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Length 364;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 123 CCCCCTGCTGGTGAAGATC 141
Db 166 CCCCCTGCTGGTGAAGATC 148

RESULT 9
US-09-703-708-10893
; Sequence 10893, Application US/09703708
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xanthomonas campestris Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15804)C
; CURRENT APPLICATION NUMBER: US/09/703,708
; CURRENT FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 1999-11-10
; PRIOR FILING DATE: US 60/163,791
; PRIOR APPLICATION NUMBER: 2000-02-22
; NUMBER OF SEQ ID NOS: 18992
; SEQ ID NO 10893
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-09-703-708-10893

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 71; Length 447;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 216 GTTGGAGCGCTGATGATC 234
Db 99 GTTGGAGCGCTGATGATC 117

RESULT 10
US-60-164-320-10893
; Sequence 10893, Application US/60164320
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xanthomonas campestris Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15804)A
; CURRENT APPLICATION NUMBER: US/60/164,320
; CURRENT FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 18992
; SEQ ID NO 10893
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Xanthomonas campestris

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 71; Length 447;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 216 GTTGGAGCGCTGATGATC 234
Db 99 GTTGGAGCGCTGATGATC 117

RESULT 11
US-60-183-791-10893
; Sequence 10893, Application US/60183791
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xanthomonas campestris Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15804)B
; CURRENT APPLICATION NUMBER: US/60/183,791
; CURRENT FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 18992
; SEQ ID NO 10893
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-60-183-791-10893

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 72; Length 447;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 216 GTTGGAGCGCTGATGATC 234
Db 99 GTTGGAGCGCTGATGATC 117

RESULT 12
US-09-270-767-11554/C
; Sequence 11554, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326 094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 11554
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11554

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Score 19; DB 18; Length 487;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 138 GATCACCCTGGAGAGGCGC 156
Db 328 GATCACCCTGGAGAGGCGC 310

RESULT 13
US-60-252-833-32151
; Sequence 32151, Application US/60252833
; GENERAL INFORMATION:
; APPLICANT: Geyer, Matthew
; APPLICANT: Geyer, Murray R.
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 105292
; CURRENT APPLICATION NUMBER: US/60/252,833
```



```

CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 43535
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 32151
LENGTH: 551
TYPE: DNA
ORGANISM: Bovine
US-60-252-833-32151
```

```

Query Match      7.2% Score 19 DB 80 Length 551
Best Local Similarity 100.0% Pred. No. 1.4e+32
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

```

CY      146 TGGACGAGGCGCCAGCCCGC 164
DB      67 TGGACGAGGCGCCAGCCCGC 85
```

```

RESULT 14
US-09-796-765-180/C
Sequence 180, Application US/09796765
GENERAL INFORMATION:
APPLICANT: Rosen, et. al.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 57
FILE REFERENCE: PO-57
CURRENT APPLICATION NUMBER: US/09/796,765
CURRENT FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/187,016
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 477
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 180
LENGTH: 572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (25)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (51)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (70)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (204)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (212)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (465)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (496)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (509)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
```

```

LOCATION: (552)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (565)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (570)
OTHER INFORMATION: n equals a,t,g, or c
US-09-796-765-180
```

```

Query Match      7.2% Score 19 DB 31 Length 572
Best Local Similarity 100.0% Pred. No. 1.4e+32
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

```

CY      25 AGCCAGGCCAGTTCACCA 43
DB      429 AGCCAGGCCAGTTCACCA 411
```

```

RESULT 15
US-09-867-682-180/C
Sequence 180, Application US/09867682
GENERAL INFORMATION:
APPLICANT: Rosen, et. al.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 57
FILE REFERENCE: PO-57C1
CURRENT APPLICATION NUMBER: US/09/867,682
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 03/796,765
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: 60/197,016
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 477
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 180
LENGTH: 572
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (26)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (51)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (53)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (170)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (204)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (212)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (465)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (496)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
```

Fri Nov 7 15:26:52 2003

us-09-622-500b-3.rtf

Page 6

```

1 LOCATION: (509)
2 OTHER INFORMATION: n equals a,t,g, or c
3 NAME/KEY: msc_feature
4 LOCATION: (552)
5 OTHER INFORMATION: n equals a,t,g, or c
6 NAME/KEY: msc_feature
7 LOCATION: (565)
8 OTHER INFORMATION: n equals a,t,g, or c
9 NAME/KEY: msc_feature
10 LOCATION: (570)
11 OTHER INFORMATION: n equals a,t,g, or c
US-09-867-682-180

```

Query Match	7.28;	Score 19;	DB 36;	Length 572;
Best Local Similarity	100.0%;	Pred. No. 1.4e+02;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 25 AGCCAGGCCCACTTCAACA 43  
 |||||:|||||:  
 Db 429 AGCCAGGCCCACTTCAACA 411

Search completed: November 7, 2003, 15:19:24  
Job time : 1869 secs

Page 6

Page 6

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:59:59 / Search time 158 Seconds  
(without alignments)  
2906.548 Million cell updates/sec

Title: US-09-622-500b-3  
Perfect score: 264  
Sequence: 1 atgacgaagtcggagatcaaa.....gcctggtccgcgcgaagtaa 264

Scoring table: OLIGO\_NJC  
Gapop 60.0, Gapext 60.0

Searched: 2459958 seqs, 869762408 residues

Word size: 0

Total number of hits satisfying chosen parameters: 49139.6

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pna/pct\_new\_comb\_seq.\*  
2: /cgn2\_6/ptodata/2/pna/us06\_new\_comb\_seq.\*  
3: /cgn2\_6/ptodata/2/pna/us07\_new\_comb\_seq.\*  
4: /cgn2\_6/ptodata/2/pna/us08\_new\_comb\_seq.\*  
5: /cgn2\_6/ptodata/2/pna/us09\_new\_comb\_seq.\*  
6: /cgn2\_6/ptodata/2/pna/us10\_new\_comb\_seq.\*  
7: /cgn2\_6/ptodata/2/pna/us60\_new\_comb\_seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	7.2	201	7	US-60-487-610-77660
2	19	7.2	201	7	US-60-485-450-39979
3	19	7.2	36530	7	US-60-487-610-19850
4	19	7.2	36530	7	US-60-485-450-12225
5	18	6.8	201	7	US-60-495-114-48489
6	18	6.8	261	5	US-09-540-2310-82306
7	18	6.8	142	6	US-10-653-595-88
8	18	6.8	183	7	US-60-507-511-4755
9	18	6.8	1928	7	US-60-507-511-4735
10	18	6.8	2115	7	US-60-507-511-2061
11	18	6.8	7800	5	US-09-674-546A-1668
12	18	6.8	42203	5	US-10-367-094-53
13	18	6.8	478484	7	US-60-495-114-16468
14	17	6.4	25	5	US-09-956-604D-512.7
15	17	6.4	25	7	US-60-507-511-44686
16	17	6.4	25	7	US-60-507-481-46080
17	17	6.4	110	6	US-10-242-535A-5290
18	17	6.4	201	7	US-60-487-610-68099
19	17	6.4	201	7	US-60-495-114-54072
20	17	6.4	201	7	US-60-500-337-117831
21	17	6.4	201	7	US-60-500-337-117858
22	17	6.4	201	7	US-60-500-337-117873
23	17	6.4	515	7	US-60-507-511-534
24	17	6.4	544	7	US-60-507-481-6553
25	17	6.4	1696	6	US-10-425-114A-14488
26	17	6.4	2043	6	US-10-425-114A-32590

C 27	17	6.4	16244	7	US-60-495-114-16510	Sequence 16510, A
C 28	17	6.4	17596	1	PCT-US01-20691.2	Sequence 2, App1
C 29	17	6.4	45824	7	US-60-500-337-19843	Sequence 19843, A
C 30	17	6.4	54464	7	US-60-495-114-16529	Sequence 16529, A
C 31	17	6.4	73256	6	US-10-332-281-167	Sequence 167, App
C 32	17	6.4	97727	7	US-60-487-610-19748	Sequence 19748, A
C 33	17	6.4	101517	7	US-60-485-450-1977	Sequence 1977, A
C 34	17	6.4	131789	7	US-60-485-114-16225	Sequence 16225, A
C 35	16	6.2	25	5	US-09-956-604D-141215	Sequence 141215, A
C 36	16	6.2	25	5	US-09-954-427A-37-071	Sequence 371071, A
C 37	16	6.1	76	1	PCT-US01-18714-20194	Sequence 20194, A
C 38	16	6.1	157	1	PCT-US01-18714-13778	Sequence 10778, A
C 39	16	6.1	194	7	US-60-507-481-9330	Sequence 9330, App
C 40	16	6.1	201	7	US-60-487-610-40597	Sequence 40597, A
C 41	16	6.1	201	7	US-60-485-114-812	Sequence 812, App
C 42	16	6.1	201	7	US-60-485-114-8125	Sequence 8125, App
C 43	16	6.1	201	7	US-60-485-114-8130	Sequence 8130, App
C 44	16	6.1	201	7	US-60-485-114-8134	Sequence 8134, App
C 45	16	6.1	201	7	US-60-485-114-8139	Sequence 8139, App

## ALIGNMENTS

RESULT 1  
US-60-487-610-77660  
Sequence 77660, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
TITLE OF INVENTION: GENETIC POLYMERISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS.  
FILE REFERENCE: CLO1469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17  
NUMBER OF SEQ ID NOS: 97101  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 77660  
LENGTH: 201  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-487-610-77660

Query Match 7.2% Score 19; Db 7; Length 201;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

27 123 CCGCGTCTGCGGAGATC 141  
28 CCGCGTCTGCGGAGATC 46

RESULT 2  
US-60-485-450-39979  
Sequence 39979, Application US/60485450  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
TITLE OF INVENTION: GENETIC POLYMERISMS ASSOCIATED WITH  
TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
FILE REFERENCE: CLO1470  
CURRENT APPLICATION NUMBER: US/60/485,450  
CURRENT FILING DATE: 2003-07-09  
NUMBER OF SEQ ID NOS: 47853  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 39979  
LENGTH: 201  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-485-450-39979



QY 127 GTGCTGTGAAGATCACC 144  
Db 179 GTGCTGTGAAGATCACC 196

## RESULT 7

US-10-653-595-86  
Sequence 88, Application US/10653595  
GENERAL INFORMATION:  
APPLICANT: Ruben et. al.  
TITLE OF INVENTION: 95 Human secreted proteins  
FILE REFERENCE: P2027P1C1  
CURRENT APPLICATION NUMBER: US/10/653,595  
CURRENT FILING DATE: 2003-09-03  
PRIOR APPLICATION NUMBER: US 09/397945  
PRIOR FILING DATE: 1999-09-17  
PRIOR APPLICATION NUMBER: PCT/US99/05804  
PRIOR FILING DATE: 1999-03-18  
PRIOR APPLICATION NUMBER: 60/078,566  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,576  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,573  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,574  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,579  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/080,314  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080,312  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/078,578  
PRIOR FILING DATE: 1998-03-19  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 470  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 88  
LENGTH: 1142  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-653-595-86

Query Match 6.8%; Score 18; DB 6; Length 1142;  
Best Local Similarity 100.0%; Pred. No. 25;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAAGATCACC 144  
Db 790 GTGCTGTGAAGATCACC 807

## RESULT 8

US-60-507-511-4755  
Sequence 4755, Application US/60507511  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Mounts, William X  
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH  
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
FILE REFERENCE: AM 101081  
CURRENT APPLICATION NUMBER: US/60/507,511  
CURRENT FILING DATE: 2003-10-02  
NUMBER OF SEQ ID NOS: 203623  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO: 4755  
LENGTH: 1823  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-507-511-4755

Query Match 6.8%; Score 18; DB 7; Length 1823;

Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 127 GTGCTGTGAAGATCACC 144  
Db 1532 GTGCTGTGAAGATCACC 1549

## RESULT 9

US-60-507-511-4135  
Sequence 4135, Application US/60507511  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Mounts, William X  
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATE  
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
FILE REFERENCE: AM 101081  
CURRENT APPLICATION NUMBER: US/60/507,511  
CURRENT FILING DATE: 2003-10-02  
NUMBER OF SEQ ID NOS: 203623  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO: 4135  
LENGTH: 1928  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-507-511-4135

Query Match 6.8%; Score 18; DB 7; Length 1928;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAAGATCACC 144  
Db 1634 GTGCTGTGAAGATCACC 1651

## RESULT 10

US-60-507-511-2061  
Sequence 2061, Application US/60507511  
GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Mounts, William X  
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATE  
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES  
FILE REFERENCE: AM 101081  
CURRENT APPLICATION NUMBER: US/60/507,511  
CURRENT FILING DATE: 2003-10-02  
NUMBER OF SEQ ID NOS: 203623  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO: 2061  
LENGTH: 2115  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-507-511-2061

Query Match 6.8%; Score 18; DB 7; Length 2115;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GTGCTGTGAAGATCACC 144  
Db 1812 GTGCTGTGAAGATCACC 1829

## RESULT 11

US-09-674-546A-1668  
Sequence 1668, Application US/09674546A  
GENERAL INFORMATION:  
APPLICANT: Institute for Genetic Research  
APPLICANT: Cation Corporation  
TITLE OF INVENTION: Neisseria Meningitidis Antigens and Compositions  
FILE REFERENCE: CHIR-0334  
CURRENT APPLICATION NUMBER: US/09/674,546A

```

CURRENT FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 3264
SOFTWARE: Patent version 3.2
SEQ ID NO 1668
LENGTH: 7800
TYPE: DNA
ORGANISM: Neisseria meningitidis
US-09-674-546A-1668

```

```

Query Match          6.8%; Score 18; DB 5; Length 7800;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      39 CACCACCCGCGAGCGCGT 55
DB      6486 CACCACCCGCGAGCGCGT 6497

```

```

RESULT 12
US-10-367-094-53/C
Sequence 53; Application US/10367094
GENERAL INFORMATION:
APPLICANT: David W. Morris
FILE REFERENCE: 529452001500
CURRENT FILING DATE: 2003-02-14
NUMBER OF SEQ ID NOS: 203
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 42203
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)....(42203)
OTHER INFORMATION: n = A,T,C or G
US-10-367-094-53

```

```

Query Match          6.8%; Score 18; DB 5; Length 42203;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      68 GCAAGCCCTACAGCCTGA 85
DB      27819 GCAAGCCCTACAGCCTGA 27802

```

```

RESULT 13
US-60-495-114-16468/C
Sequence 16468; Application US/60495114
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
FILE REFERENCE: CL001480
CURRENT FILING DATE: 2003-08-15
NUMBER OF SEQ ID NOS: 91238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16468
LENGTH: 478484
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)....(478484)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-495-114-16468

```

```

Query Match          6.8%; Score 15; DB 7; Length 478484;

```

```

Best Local Similarity 100.0%; Pred. No. 19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      121 TGGTGAAGATCACTGCG 148
DB      328591 TGGTGAAGATCACTGCG 328674

```

```

RESULT 14
US-09-396-604D-51217
Sequence 51217; Application US/0996604D
GENERAL INFORMATION:
APPLICANT: Miltmann, Michael
TITLE OF INVENTION: Methods of Genetic Analysis of Escherichia coli
FILE REFERENCE: 3117.1
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/234,049
CURRENT FILING DATE: 2000-09-19
NUMBER OF SEQ ID NOS: 14629
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.2
SEQ ID NO 51217
LENGTH: 25
TYPE: DNA
ORGANISM: E. coli
US-09-396-604D-51217

```

```

Query Match          6.4%; Score 17; DB 5; Length 25;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      212 GCCAGTTCGGCAGCCTG 228
DB      4 GCCAGTTCGGCAGCCTG 26

```

```

RESULT 15
US-60-507-511-44688
Sequence 44688; Application US/60507511
GENERAL INFORMATION:
APPLICANT: Mounts, William W
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATE
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
FILE REFERENCE: AM 101081
CURRENT FILING DATE: 2003-10-02
NUMBER OF SEQ ID NOS: 203623
SOFTWARE: Patent version 3.2
SEQ ID NO 44688
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapiens
US-60-507-511-44688

```

```

Query Match          6.4%; Score 17; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY      8 AGGTGAATCAACGCC 24
DB      5 AGGTGAATCAACGCC 21

```

```

Search completed: November 7, 2003, 15:22:11
Job time : 161 secs

```



ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11509

Query Match 7.2% Score 19; DB 4; Length 1872;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CGTGGTGGTGAAGATCACC 144  
DB 1143 CGTGGTGGTGAAGATCACC 161

## RESULT 3

US-09-252-991A-3380/C  
Sequence 3380, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 3380  
LENGTH: 1923  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3380

Query Match 7.2% Score 19; DB 4; Length 1923;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CGTGGTGGTGAAGATCACC 144  
DB 1027 CGTGGTGGTGAAGATCACC 1009

## RESULT 4

US-09-252-991A-11509  
Sequence 11509, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11509  
LENGTH: 456  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11509

Query Match 6.8% Score 18; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 GCGTGAATGCAACCGCC 241  
DB 167, GCGTGAATGCAACCGCC 184

RESULT 5  
US-09-252-991A-11732/C  
Sequence 11732, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11732  
LENGTH: 1059  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11732

Query Match 6.8% Score 18; DB 4; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 GCGTGAATGCAACCGCC 241  
DB 657 GCGTGAATGCAACCGCC 843

## RESULT 6

US-09-252-991A-11984/C  
Sequence 11984, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-19  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11984  
LENGTH: 1111  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11984

Query Match 6.8% Score 18; DB 4; Length 1111;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TCGGCACTGATGATCG 235  
DB 152 TCGGCACTGATGATCG 135

## RESULT 7

US-09-252-991A-12028  
Sequence 12028, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A



;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 12028  
;; LENGTH: 1797  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12028

Query Match 6.8%; Score 18; DB 4; Length 1797;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TCGGAGCCTGATGTCG 235  
Db 1652 TCGGAGCCTGATGTCG 1669

RESULT 9  
US-09-252-991A-11628  
;; Sequence 1628, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 11628  
;; LENGTH: 2136  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11628

Query Match 6.8%; Score 18; DB 4; Length 2136;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 GCGTGATGTCGACGGCC 241  
Db 453 GCGTGATGTCGACGGCC 470

RESULT 9  
US-09-252-991A-11913/c  
;; Sequence 11913, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 11913  
;; LENGTH: 2334  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11913

Query Match 6.8%; Score 18; DB 4; Length 2334;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TCGGAGCCTGATGTCG 235  
Db 2248 TCGGAGCCTGATGTCG 2231

RESULT 10  
US-09-252-991A-11999  
;; Sequence 11999, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 11999  
;; LENGTH: 2436  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11999

Query Match 6.8%; Score 18; DB 4; Length 2436;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TCGGAGCCTGATGTCG 235  
Db 1455 TCGGAGCCTGATGTCG 1472

RESULT 11  
US-08-459-586-5  
;; Sequence 5, Application US/08459586  
;; Patent No. 5720957  
;; GENERAL INFORMATION:  
;; APPLICANT: Jones, Thomas R.  
;; TITLE OF INVENTION: Identification of a Human  
;; TITLE OF INVENTION: Cytomegalovirus Gene Region Involved in Down-Regulation  
;; NUMBER OF SEQUENCES: 20  
;; CORRESPONDENCE ADDRESSES:  
;; ADDRESSEE: American Cyanamid Company  
;; STREET: One Campus Drive  
;; CITY: Parsippany  
;; STATE: New Jersey  
;; COUNTRY: United States  
;; ZIP: 07054  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.10  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/459,586  
;; FILING DATE: 02-03-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Barnhard, Elizabeth M.  
;; REGISTRATION NUMBER: 31,088  
;; REFERENCE/DOCKET NUMBER: 32,425-01  
;; TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-459-586-5

Query Match 6.8%; Score 18; DB 1; Length 5020;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ACAGTACCCCGCTGG 133  
|||||  
DB 2593 ACAGTACCCCGCTGG 2610

## RESULT 12

US-08-282-696-5  
Sequence 5: Application US/08282696  
Patent No. 5846936  
GENERAL INFORMATION:  
APPLICANT: Jones, Thomas R.  
APPLICANT: Campbell, Ann E.  
TITLE OF INVENTION: Identification of a Human  
TITLE OF INVENTION: Cytoomega-cytus Gene Region Involved in Down-Regulation of  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07470-8426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/282,696  
FILING DATE: 23-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,425-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3246  
TELEFAX: 201-831-3305  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-282-696-5

Query Match 6.8%; Score 18; DB 2; Length 5020;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ACAGTACCCCGCTGG 133  
|||||  
DB 2593 ACAGTACCCCGCTGG 2610

## RESULT 13

US-08-459-586-3  
Sequence 3: Application US/08459586  
Patent No. 5720957  
GENERAL INFORMATION:  
APPLICANT: Jones, Thomas R.  
APPLICANT: Campbell, Ann E.  
TITLE OF INVENTION: Identification of a Human  
TITLE OF INVENTION: Cytoomega-cytus Gene Region Involved in Down-Regulation  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Campus Drive  
CITY: Parsippany  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,586  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,425-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7242 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-459-586-3

Query Match 6.8%; Score 18; DB 1; Length 7242;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ACAGTACCCCGCTGG 133  
|||||  
DB 5311 ACAGTACCCCGCTGG 5328

RESULT 14  
US-08-282-696-3  
Sequence 3: Application US/08282696  
Patent No. 5846936  
GENERAL INFORMATION:  
APPLICANT: Jones, Thomas R.  
APPLICANT: Campbell, Ann E.  
TITLE OF INVENTION: Identification of a Human  
TITLE OF INVENTION: Cytoomega-cytus Gene Region Involved in Down-Regulation  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07470-8426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/282,696  
FILING DATE: 23-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 32,425-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3246  
TELEFAX: 201-831-3305  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-282-696-3

Query Match 6.8%; Score 18; DB 2; Length 5020;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/262,696
FILING DATE: 29-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 33,088
REFERENCE/DOCKET NUMBER: 32,425-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3246
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 7242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-282-696-3

```

```

Query Match          6.8%; Score 18; DB 2; Length 7242;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      116 ACAGTACCCCGTGGTGG 133
      |||||..|||..|||
Db      5311 ACAGTACCCCGTGGTGG 5328

```

```

RESULT 15
US-09-103-840A-2
Sequence 2, Application US/09/03840A
Parent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert J.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-2007,00
CURRENT APPLICATION NUMBER: US/09/03,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

Query Match          6.8%; Score 18; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 7,47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      41 CCACCCGACGGCGCTGA 58
      |||||..|||..|||
Db      767498 CCACCCGACGGCGCTGA 767515

```

Search completed: November 7, 2003, 14:48:16  
 Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw mode.

Run on: November 7, 2003, 12:27:20 (Search time 1599 seconds)

(without alignments)  
7157.151 Million cell updates/sec

Title: US-09-622-500b-3

Perfect score: 264

Sequence: 1 atgacaaagcggagatcaaa.....gcccgtgcccgcgaagtaa 264

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2880721 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

GenEmbl:  
1: gb\_ba:  
2: gb\_hg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pac:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vl:  
15: em\_ba:  
16: ex\_fun:  
17: em\_hum:  
18: ex\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_ot:  
22: em\_ov:  
23: em\_pat:  
24: em\_ph:  
25: em\_pl:  
26: em\_ro:  
27: em\_sts:  
28: em\_un:  
29: ex\_vl:  
30: em\_hg\_hum:  
31: em\_hg\_inv:  
32: em\_hg\_other:  
33: em\_hg\_mus:  
34: em\_hg\_pun:  
35: em\_hg\_rtd:  
36: em\_hg\_mam:  
37: em\_hg\_vrt:  
38: em\_sy:  
39: em\_hggo\_hum:  
40: em\_hggo\_mus:  
41: em\_hggo\_other:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147.2	55.8	6407	7	FI0CG
2	147.2	55.8	6407	7	INXPF
3	147.2	55.8	6971	6	AX006030
4	147.2	55.8	6971	6	BD131:575
5	145.6	55.2	6408	7	INPDX
6	145.6	55.2	6408	7	PPDCG
7	145.6	55.2	7055	6	AX006017
8	145.6	55.2	7055	6	BD131:874
9	145.6	55.2	7775	12	AF464139
10	145.6	55.2	7783	6	AX006032
11	145.6	55.2	7783	6	BD131:871
12	145.6	55.2	8233	12	AF362081
13	145.6	55.2	8879	12	AF218733
14	145.6	55.2	9183	12	AF217317
15	145.6	55.2	9183	12	AF218734
16	145.6	55.2	9183	12	AF464138
17	145.6	55.2	9198	12	AF246447
18	145.6	55.2	9206	12	AF218364
19	145.6	55.2	9225	12	AF246446
20	145.6	55.2	9234	12	AF218353
21	145.6	55.2	9252	12	AF246445
22	145.6	55.2	9264	12	AF246446
23	145.6	55.2	9267	12	AF246449
24	145.6	55.2	9270	12	AF246450
25	145.6	55.2	9273	12	AF246448
26	145.6	55.2	9273	12	AF246451
27	145.6	55.2	9273	12	AF246452
28	145.6	55.2	9273	12	AF246453
29	145.6	55.2	9276	12	AF246454
30	145.6	55.2	9279	12	AF246455
31	144	54.5	716	7	FI0G578
32	144	54.5	6437	7	INX13X
33	144	54.5	6867	12	SYNMBE23V
34	144	54.5	6871	12	SYNMBE2342
35	144	54.5	6871	12	SYNMBE2341
36	144	54.5	6940	12	SYNMBE2344
37	144	54.5	7083	6	AR13:047
38	144	54.5	7083	6	AR261:939
39	144	54.5	7196	12	SYNMBE3MD1V
40	144	54.5	7196	12	SYNMBE3MD2V
41	144	54.5	7229	12	SYNMBE3MD8V
42	144	54.5	7229	12	SYNMBE3MD8V
43	144	54.5	7232	12	SYNMBE3MD8V
44	144	54.5	7238	12	SYNMBE3MD8V
45	144	54.5	7244	12	SYNMBE3MD8V

## ALIGNMENTS

RESULT 1  
LOCUS FI0CG  
DEFINITION Bacteriophage fi, complete genome.  
ACCESSION U02448.1  
VERSION U02448.1 GI:166201  
KEYWORDS complete genome; origin of replication.  
SOURCE Enterobacteriophage fi  
ORGANISM Enterobacteriophage fi  
REFERENCE 1. Jhaeas 1 to 6407.  
AUTHORS H.L.D.F. and Petersen.G.B.  
TITLE Nucleotide sequence of bacteriophage fi DNA  
JOURNAL J. Virol. 44 (1), 32-46 (1982)  
MEDLINE 83059852

Pred. No. is the number of results predicted by chance to have a

**FJ06MED**      6292494  
**COMMENT**    Original  
**FEATURES**

        source  
             /organism="Enterobacteria phage f1"  
             location/Qualifiers  
             1..6407  
             /mol\_type="genomic DNA"  
             /db\_xref="taxon:10863"  
             join(6006..6407)..83-)

CDS  
             note="protein I1"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32209.1"  
             /db\_xref="GI:166202"  
             /translation="MIDMLVLRCPFDISVCSSGSDQCLAFVDLSKATITSGNLSA  
RTVEHTIGDLYVGSLSHPFESLPTHYSTARRYEGSNFPPCEITASAKYLCH  
NWFGCTDALCSEALKLPFNSFCLELDLYNAATTISKDAFTSRAPENLNKYVI  
DLRWNSNGOTKSIRSONWKESTVTWNESHRLAVLKGVLCIQOIQQLSKPSAMT  
TSYQEOLKTVINSPDLLFEFAGLVREEARIKTYLKSGLPLNLPCA-RPASYNRSYG  
KDILJDLMVSFSFSAFLPAEEDMNIVDSAVALDIQSKEPFTTPSCITSASKARYE  
GFRIALIVEGYDSVALTMPRNSFWRYALSALVEGGIPKSQLMLNSTGNVVPLVFNVDFSSGRPDW  
DFSSGRPDWNEVPVKLIA"

        exon  
             1..>831  
             /note="protein II"  
             496..831  
             /note="protein X"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32210.1"  
             /db\_xref="GI:166203"  
             /translation="MWNYVDSALVLDAIQQKHFTTPSQKTSFAKASYRGYSRRUNE  
GYDSVALTMPRNSFWRYALSALVEGGIPKSQLMLNSTGNVVPLVFNVDFSSGRPDW  
INEPVKLIA"

CDS  
             843..1106  
             /note="protein V"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32211.1"  
             /db\_xref="GI:166204"  
             /translation="MKVIEIKPSQAQFTTRSGVSRQKPYSINEQCYVDLGNEYPC  
VKITDEGPAYAAGLTVHLLSFYKGCFGSIMDRKLVPAK"

CDS  
             1108..1203  
             /note="protein VII"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32212.1"  
             /db\_xref="GI:166205"  
             /translation="IMEGWADBDTIYOAMQISYVLCFALGITAGCOR"

CDS  
             1206..1304  
             /note="protein IX"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32213.1"  
             /db\_xref="GI:166206"  
             /translation="MSVLYVVSFPASFVLCMRBSGTTFPRIMETSS"

CDS  
             1301..1522  
             /note="protein VIII"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32214.1"  
             /db\_xref="GI:166207"  
             /translation="MKRSLYKASVAVALTVPMLSFAEGDDPAFAAFCSACASEATEY  
IGYAAMVWVVGATIGIXLFKKFTSKAS"

CDS  
             1579..2851  
             /note="protein III"  
             /codon\_start=1  
             /transl\_table=1  
             /protein\_id="AAA32215.1"  
             /db\_xref="GI:166208"  
             /translation="IMKKCLPAIPLVVPFYSHASAEVESCLAKPHENSFTMMKDCXET  
LBRAVAGCCLMNTGVVNCCTGETOOQGWWPIGLAIPENGSGSGSGSGSGSGSG  
GGTKPEPGGTRPGYTVINDPLDGTYPPGTBNPNPNLSLESPLMTFMFOKNRR  
NRQGLALTTCVTYQTGDTPKITYQTTPVSSKAMDTAINMGFRCDARHGSGFNDFPV  
CEIQQSQSDLLQPVPNAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG

```

DYMALNATCAAGTETWADENALDQSDAKGKLDIVATGGAIDEEAGVDSIANGNGNAT
GDGASNSGMAAGVGGDGNNSLDENLNNFPOYPSIQQNVCECPPTVGAGKAPYEFSSIDQK
NTRGVPAPFLIYATYATYVYPSITANILRNKNS"
2856..3184
/note="protein v"
/codon_start=1
/translation="
/protein_id="AAA32216.1"
/db_xref="gi:166209"
/translation="MRYLGIIRLLRQSLITVTPYKLIITFLKKGGKILAIASLEL
ALILGSLIDVGTLSAQJPSFVQGVCLITLPSKLLPQYVLSVPAIFITVKK
KIVSYLDWCK"
3186..4342
/note="protein v"
/codon_start=1
/translation="
/protein_id="AAA32217.1"
/db_xref="gi:166210"
/translation="MAVYVIGKLDGSGKTVSYNGKIDKKVANGKILATYKLIPLQNP
QVRFAPKTPVRLIPKPSISLALIGRNSHSDENNLIVLDECGTFEYRSMK
ERGLIITWELMARKLWELIFVLECLIVKQKASALANVYQERLQRLTPVCT
VSLTSCGKSPKLIHGVVYGEISLSETPYEPNLYTKNNYVADTQAFSSNDYK
VSLTIPSLSKPPEKPEKPNISQKKITKTYIKKPEVCLAIIPASATYVITTPKPE
VKKYSGTTPPEKFTIDSSORLNLSTYVFKDQKKLNSDLDQKQSLYITLCLTV
SIRKGVSNELVKN"
4342..4500
/note="protein v"
/codon_start=1
/translation="
/translation="MKLIATINPFLKPSSSSSAQVIEKNNSLRLPVTYKSGTGE
SVLSPVKGITVTVSSDVKPEWLRNFISVLRRNPFDMVGSNSIILQKVPNNQDIT
DELPSSNDGSDNSAPSGGFPPQDNQVOTKINNRKALILRVLEFVKSITSGS
SNVLSIGSNLIVSAPKQILDNPCPLSTVLLPTDQILLEGILFEVQGGALDFEPA
AGSGRGVAGGANNTRTSVLSVSSAGSEEGFNGVGLSLPALKTSNHSKLIASPRIL
TSGOKGSLVGNQNPETTSVIVGSESNVNNPQCTIRQVGLISKSYVPMAGSNTV
LDITRKADSLSSTQASDVIITKRSITATYNLSDGCLLDGSLTFPKNTSDSSVPLP
SKPLIGLPSFSSSDSNEESITVYLVKATIVRAL"
4500..6407
/exon
/note="protein 1"
BASE COUNT 1574 a 1298 c 1325 g 2210 t
ORIGIN
Query Match 55.8%; Score 147.2; DB 7; Length 6407;
Best Local Similarity 72.3%; Pred. No. 7.9e-14;
Matches 191; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
CY 1 ATGATCAAGTGGGATGATCAAGCCGAGCAGGCGCCACTTCCACCGCAGCGCGCTAGC 60
Db 543 ATGATTTAAAGTTAAATTAACCACTCTCAAGCGCAATTCCTACCCGCTTCTG337TCT 902
CY 61 CGCCAGGCGCAAGCCCTTCAAGCCTGACGACGACGCTGTGCTAAGTGAAGCTGG39CAAG 120
Db 303 CGTCA39GCAAGCCTTACTTCATGAAAGACGACGACGCTTGTACGTTAATTGG37TAACGA 362
CY 121 TACCCCGTGGCGGAGATACACCTGGGACGACGACGACGACGACGCTGCGTCCGCGCGCTG 180
Db 963 TATCCGATGCTGTCAAGATTAATTTGTGATGAAGGATGAGGACGACGCTATCGCGCTGGTGT 10222
CY 181 TACACCTGACACCTGACGACGCTTCAAGGTGGGACGACGACGACGACGCTGATGATGACGCG 240
Db 1223 TACACCGTTCACTGTCTGTCTGTAAGATTCGCTGATGATGATGATGATGATGATGATGAT 1882
CY 242 CTGCGGCTTGCTGCCCGCGCAAGTAA 284
Db 1083 CTGCGGCTTGCTGCCCGCGCAAGTAA 1106

```

```

ACCESSION      V00606.1 J02449 M10881
VERSION        V00606.1 GI:14974
KEYWORDS       circular; coat protein; genome; origin of replication.
SOURCE         Enterobacteria phage f1
ORGANISM       Enterobacteria phage f1
REFERENCE      1 (bases 1 to 6407);
AUTHORS        Beck, S. and Zink, S.
TITLE          Nucleotide sequence and genome organization of filamentous
JOURNAL        Gene 16 (1-3), 35-58 (1981)
MEDLINE        82211801
PUBMED         6282703
REFERENCE      2 (bases 5493 to 5852)
AUTHORS        La Farina, M. and Vitale, M.
TITLE          Rio-dependence of the terminator active at the end of the I region
JOURNAL        Mol. Gen. Genet. 195 (1-2), 5-9 (1984)
MEDLINE        85035856
PUBMED         6092864

FEATURES
     source          1..6407
                     /organism="Enterobacteria phage f1"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10863"
                     join(6006..6407,1..831)
                     /gene="ii"
                     join(6006..6407,1..831)
                     /gene="ii"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA23876.1"
                     /db_xref="GI:603850"
                     /translation="MIDMLVLRPFTLSVCSRSLSGNDLIAVFDISKATISGKLSA
                     RVEHFDGDLTVSGLSHPESLPHYSGLAKITGSKNRPPEVETASFAKVCQGH
                     NWFGTDLACSEALLNFANSLPCLYDLADNATISRTIDTFSAKAPNENIAKQVI
                     DLRANVNGQTKSRQSNWSTVTWNETRHRITLVAALKHEVLOHQIQQLSSKSAAN
                     KSYQKQKLVLSNPDLSEFASGLVREARLETVYKSKFGLPLNLDARFASDYNQ3
                     KD.IFDLMSFSELPKAFEGSGSNLYDSAVLDA.QSGHPTITPSGKTSFAKSAKRY
                     GFYRLPNEGYSVAL.TMPRNSFMRYVSALVSCGI.FKSQLMNSTCNVPLVNFIV
                     DESSORPDWNEPVAKIA"
                     496..831
                     /note="gene X"
                     /gene="X"
                     /codon_start=1
                     /trans_table=11
                     /protein_id="CAA23867.1"
                     /db_xref="GI:14975"
                     /db_xref="SWISS-PROT:P03659"
                     /translation="MKNVYDSAVLDAIQSKFRTTPSGKTSFAKSAKRYGFRRLAKK
                     GYDSVALTMPRNSFMRYVSALVSCGIFKSQLMNSTCNVPLVNFIVDESSQSPDW
                     VNEPVAKIA"
                     843..1106
                     /note="gene V"
                     /codon_start=1
                     /trans_table=11
                     /protein_id="CAA23868.1"
                     /db_xref="GI:14976"
                     /db_xref="SWISS-PROT:P03669"
                     /translation="MKNVEIKPSQAQFTTSGVSRQSKPYSLNEQLGVYDLGNEYVVL
                     VKITLDEGPAYAPGLYTHLSFVKGQGS.LNDRLRVPAK"
                     1108..1209
                     /note="gene VII"
                     /codon_start=1
                     /trans_table=11
                     /protein_id="CAA23869.1"
                     /db_xref="GI:14977"
                     /db_xref="SWISS-PROT:P03675"
                     /translation="MEGVADFCITIQAMQISVLCFALGIIAGGQR"
                     1206..1304
                     /note="gene IX"
                     /codon_start=1

CDS
     1..6407
     /trans_table=11
     /protein_id="CAA23870.1"
     /db_xref="GI:14978"
     /db_xref="SWISS-PROT:P03677"
     /translation="VSVLVSFASFAFV.GWCLRSQTYFTFLWEFS"
     131..1522
     /note="gene VII"
     /codon_start=1
     /trans_table=11
     /protein_id="CAA23871.1"
     /db_xref="GI:14979"
     /db_xref="SWISS-PROT:P03617"
     /translation="MKKSLVAKSAVAVAT.VPMLSFAAEGCDFAKAAFDLSQASATEY
     IGHANVYVNGATG.KLPKRTSAS"
     1579..1853
     /note="gene II"
     /codon_start=1
     /trans_table=11
     /protein_id="CAA23872.1"
     /db_xref="GI:579103"
     /db_xref="SWISS-PROT:P03662"
     /translation="MKKLIFAIPVWPVYSHAFIVSGCAKPTENSFTNWKMDYT
     LDRANVSGCLMA.GVVVCTGDEFCQGWPIQLA1PEYEGGSGSEGGEGGSGS
     GGTXPPEYQDPIPGYTIINPLDGYTPGTQNPANLPSLESOP.NTFMFKNNRER
     NRGALTYTGTCGCTDQPVKTYOYTVSSKAMVDAMNGKPRDCAHSGFNDPBY
     CYGGQSSDLPPQPVNAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGS
     DYENMANANKAKT.EKADENLQSDAKGKIDVAITDGAALDGTGYSGLANNGAT
     GFASNSQMAQVGGSDNSPLMNNFQYLPSPQSVBGRPVFRAGKPYRSTCCXI
     N.FRGVFAFLVAVTFYVFTFANLKNKS"
     2856..3194
     /note="gene VI"
     /codon_start=1
     /trans_table=11
     /protein_id="CAA23873.1"
     /db_xref="GI:14981"
     /db_xref="SWISS-PROT:P03673"
     /translation="VYVLAGPLDLRFGLPLVTCDFGLLTFJAKGRKQIAIA.SFL
     A.IGLNLSLVYSLDISACHPSPVQGVQV.IPSNALPCYVTLVYAAFIIDWK
     KVSALENKX"
     3196..4242
     /note="gene I"
     /codon_start=1
     /trans_table=11
     /protein_id="CAA23974.1"
     /db_xref="GI:14982"
     /db_xref="SWISS-PROT:P03657"
     /translation="MAVYEVGKSGSEKTLVSK.CKRIYASCKIATNLDPLQNDP
     OGRPAKTRPRVRI.PKPSISDILA.GSGNDSYDENKGLVLECGKWKTPBEMDK
     BKQIITLFLHARKKQK.IFVQCLSTVDQNSALANRYVCEKRRITIPVSTL
     YSLITGSKPCLKHLGVVKKIGSLSTPVERVYTSKLVNADTGRHSSNTDSGV
     VSLTFYLSHGRFEKIVLQGVKYLTVLKKFSEVCLAGFAAFYVSYITPKRE
     VKKQSVTDEPDKRT.DSSQNLNLYEVFKSKKGLNSDLOQGYSLYIDLCYV
     SIKKQSNELVKCN"
     4220..5800
     /note="gene IV"
     /codon_start=1
     /trans_table=11
     /protein_id="CAA23875.1"
     /db_xref="GI:14983"
     /db_xref="SWISS-PROT:P03666"
     /translation="MKLVNVEFLMNVSSSSFAQVTEMMNSGJCFVTVYSKQTE
     SVUSPDVQGVTVYVSSVYKPNLEPFTVYPRNPEMWSISITICKVYANNQDVI
     DE.PSSNDQEPDKNAPSGSGGFVPPONKAVTQFKNNVRAKCLRIVEG.FPKSNTSKS
     SVLSDVGSNLLVSAFKQIDNLDPQLSTVDLPLDQILIEGLIFEVQGCALDRSPA
     AASORGTVAGVNTDRILTVLSAGSGGIFNGVGLSVALKTNSSKLSVPRIL
     TSGQKGSISVQGNVPFITGRVTCGSANVKNPCTVEICNQCISMSVDPVAGGNIV
     LDTSKADS.SSTOASDQSVITNORSIATNNPDQOTLLGLITVYKVTSGSGVFFL
     SKIPLIGLFSRSRSDNEESTLYVIVKATIVAA"

BASE COUNT      1569 a 1298 c 1328 g 2212 t
ORIGIN
Query Match      55.8%; Score 147.2; DB 7; Length 6407;
```

```

/transl_table=11
/protein_id="CAC05235.1"
/db_xref="cg:5928969"
/translation="MKSLVAKAVAVATVPMSPFAEGDPKAAFDSDHDSNTFY
GYAAVVVVGATGIGKIKKFTSKAS"
5289..5948

```





```
Db 4609 TATCCGGTGTTCGAAGATACCTCTTGATGAAGAGGCGACGACCTATCGGCTGTG 4668
181 TACACCGTGACACCTGACAGAGCTTCAAGGTGGGCGAGTGGAGCTGATGATCGACCGC 240
Db 4669 TACACCGTGACACCTGCTCTGCTTCAAGAGTGGGCGAGTGGAGCTGATGATGACCGT 4728
Cy 241 CTGCGCTGTGTGCGCGCAAGTAA 264
Db 4729 CTGCGCTGTGTGCGCGCAAGTAA 4752

RESULT 5
INFO
LOCUS 4408 bp DNA circular PG3-10-FEB-1993
DEFINITION Genome of the bacteriophage fd (Inoviridae).
ACCESSION V00632 J02451 M10731 M10787 M21665 M21667 M21668 M21669 M21670
M21698
VERSION V00632.1 GI:14931
KEYWORDS circular; coat protein; genome; origin of replication.
SOURCE Enterobacteria phage fd
ORGANISM Enterobacteria phage fd
Virus; ssDNA viruses; Inoviridae; Inovirus.
1 (bases 1 to 6408)
REFERENCE Beck, E., Sommer, R., Auerwald, E. A., Kurz, C., Zink, B., Osterburg, G.,
Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M.
JOURNAL Nucleotide sequence of bacteriophage fd DNA
MEDLINE Nucleic Acids Res. 5 (12), 4495-4503 (1978)
79:36480
COMMENT KST FD
745987

FEATURES
source
Note difference from author's enumeration.
Location/Qualifiers
1..6408
/organism="Enterobacteria phage fd"
/mol_type="genomic DNA"
/db_xref="taxon:10864"
226..1458
/feature="GENE 1"
/codon_start=1
/transl_table=1
/protein_id="CAA23845.1"
/db_xref="GI:14932"
/db_xref="SWISS-PROT:P03659"
/transl_table="MIDMLYRLPFIQSVCQSRISQNDIAFVLSKNTJSGMNLISA
RTVEYHIDGDVSGSHPPESPLHYSGIAFKIYEGSKKPYCVAIKASPAKLVQGH
NFGCTDLACSEALLENFANSLPCVLDVAVATISRIDATFSARAPENITACVI
DHLBVNSGOTKSTRSQMESTVTMTNSTRILVAVKHVELGHOIQGLSKSPAKM
TSYOKDOLKUSNPDLEPAGVLRPEARIEETVYKSGFGLPNTLPDAITPASCYNOC
KDLIFDLMSFSFSELPKAFSGDDMKITDCAVUDAOSKHFTITPBGKTSFKAARIF
GFRRLVNEGDSVLTMPNRSFMRVSAVLEGIFRSQMLNLSICNNVPLVRLIV
DFSSQRPDYNPVLKIA"
1123..1458
/feature="GENE X"
/codon_start=1
/transl_table=1
/protein_id="CAA23846.1"
/db_xref="GI:14933"
/db_xref="SWISS-PROT:P03659"
/transl_table="YN:YDQSAVUDAOSKHFTITPBGKTSFKAARIFGFRRLVNE
GYDSVALTMPNRSFMRVSAVLEGIFRSQMLNLSICNNVPLVRLIVDFSSQRPDW
YNEPVLKIA"
1470..1733
/feature="GENE V"
/codon_start=1
/transl_table=1
/protein_id="CAA23847.1"
/db_xref="GI:14934"
/db_xref="SWISS-PROT:P03669"
/transl_table="MI:KVELIKPQQAOTSSGVSQCKEYSLNEOCYVDLNGYFVL
VKITLDSQAPYAPGLVTHSSFKVSGFOSLMDRLRLVPAK"
1735..1836
/feature="GENE VII"
/codon_start=1
/transl_table=1
/protein_id="CAA23848.1"
/db_xref="GI:14935"
/db_xref="SWISS-PROT:P03675"
/transl_table="MEQVAFPTIYQAMIQSVLCEPALG-IACGQR"
1833..1931
/feature="GENE IX"
/codon_start=1
/transl_table=1
/protein_id="CAA23849.1"
/db_xref="GI:14936"
/db_xref="SWISS-PROT:P03677"
/transl_table="MSVLYGFSASVLMQCKRSGLTYSRLKETS"
1929..2149
/feature="GENE VIII"
/codon_start=1
/transl_table=1
/protein_id="CAA23850.1"
/db_xref="GI:14937"
/db_xref="SWISS-PROT:P03678"
/transl_table="MKSLFAFLVVFYRSSEAEVSECLAXPHTENSFTWAKTKT
LDRVANYEGCLNNAQVWICGICDTCQCTQCTWTFIGIAIPENSGGSGSGSGSG
GGTRPEPEGDPPIGVYVINTPLDGTGFTGEQNPANPWSLEESQPLTFPQNRRF
NROGALTIVTGTGTCIDPKYVYTPYSKAYDAVWNSKFRCAHPSSENPFPV
CEGCGSSSLPQPPVMAAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
DYEMANKAKATENAPENALOSPAAKGLDVAITDGAITGIFGDSGLANGCAT
GDFAGNSQMAVQGDNSPLMNTROYPSLPQSVBQRPVFGAGKPYRFSIDCKI
NFRGVFAFLVYATFVYVFTFANLRKES"
3483..3821
/feature="GENE VI"
/codon_start=1
/transl_table=1
/protein_id="CAA23852.1"
/db_xref="GI:14939"
/db_xref="SWISS-PROT:P03673"
/transl_table="YEV:IGPLDIFRFLQFLVCTFGVLTPELKGFGKIA-AISPL
ALIGNSILVGRSLQISQLPDSDFQGVQCTLPSPALPCFVYLISVAIFIFDVQ
KISYLDWIK"
3824..4870
/feature="GENE I"
/codon_start=1
/transl_table=1
/protein_id="CAA23853.1"
/db_xref="GI:14940"
/db_xref="SWISS-PROT:P03655"
/transl_table="YAVVYVTRGKISGKIVSGKIQKIVAGCKLATYD-RILNLP
QNGPEAKTPVLAIRPKPSLPLATGNDSDYDKNKGLVLEDEGNTERRSNXK
BROPIIDWEHAKLGMDFLYDQISYDRARASALRHVYQRLDQNTTPPVTU
YSVLTGSKVLPKLVHGVGVYQDSQSLPVEBALYGAJNNAVITKQKSSVYQSG
YSVLTPLYSQGRFRPJAAGQVKKIKIKYCKFSRLCAIFASAFVYSITQKPE
VKRVQGTQDPDFTIDSSQRLNLSYRVFKDSKSKLIMEDLQKQGSITVTDCTV
SIKXNSNEVKN"
4848..6128
/feature="GENE IV"
/codon_start=1
/transl_table=1
/protein_id="CAA23854.1"
/db_xref="GI:14941"
/db_xref="SWISS-PROT:P03664"
/transl_table="MKL:LVIRVFLMFVSSSSFAQVLEWNSPLRQVTVWYKQGE
SVYVSDVQVGTIVYSQDVYKQENLRFFSYLRANFDQVSGSPATIQKYNRSDQYI
DELPSDDIODEYDNSAPSGGFFVPQNDVNTCTFKINNVAKDLIRIVEFVKSNSKS
SNVLSDGSLNLVNSAPKIDNLDPQFSTVDLPIDQILIEGLIFEVQGDALDFSA
```

AGSOGCTVAGGNTBRLTISVSSAGGSFGIENGEDVLSVFLAKTNHSKRLSVERIL  
 TSSGKGISIVGQNVFIFIGRTVGTGSANVNNPFCITVROVNGISNVFPPVMAAGNIV  
 LDTISVADSLSSTQSDVITNORSIACTVN-RCSQTLILGLTDVKNYSCSSVPP  
 SKPLIGLIFSSRSDSNESTIYLVKATIVALL"

BASE COUNT 1578 a 1285 c 1325 g 2210 t

ORIGIN

Query Match 55.2%; Score 145.6; DB 7; Length 6408;  
 Best Local Similarity 72.0%; Pred. No. 1,4e-13;  
 Matches 100; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGATCAAGCTGAGATCAAGCCGACGCCCGCTTACACACCCGAGCGGCTTACG 60  
 DB 1470 ATGATCAAGCTGAGATCAAGCCGACGCCCGCTTACACACCCGAGCGGCTTACG 1529  
 QY 61 CGGAGGCGAGCGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 120  
 DB 1530 CGGAGGCGAGCGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 1559  
 QY 121 TACCGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 180  
 DB 1590 TACCGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 1649  
 QY 181 TACACCGTGAAGCTGAGCGGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 240  
 DB 1650 TACACCGTGAAGCTGAGCGGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 1705  
 QY 241 CTGCGGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 264  
 DB 1710 CTGCGGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 1733

RESULT 6  
 LOCUS PFDCG 6408 bp ss-DNA circular PHG 28-Apr-1993  
 DEFINITION Bacteriophage fd, strain 478, complete genome  
 ACCESSION J02451 M10731 M10731 M21666 M21667 M21668 M21669 M21670 M21671  
 VERSION J02451.1 GI:215394  
 KEYWORDS DNA-binding protein; circular; coat protein; complete genome;  
 origin of replication; unidentified reading frame;  
 Enterobacteria phage fd  
 Enterobacteria phage fd  
 Viruses; ssDNA viruses; Inoviridae; Inovirus.

REFERENCE 1 (sites)  
 Sugimoto, K., Sugisaki, H., Okamoto, T. and Takamami, M.  
 Studies on bacteriophage fd DNA. II. Nucleotide sequence preceding  
 the RNA start-site on a promoter-containing fragment  
 Nucleic Acids Res. 2 (11), 2091-2100 (1975)  
 78248318  
 JOURNAL MEDLINE  
 PUBMED 1052531  
 AUTHORS Schaller, H., Gray, C. and Herrmann, K.  
 Nucleotide sequence of an RNA polymerase binding site from the DNA  
 of bacteriophage fd  
 Proc. Natl. Acad. Sci. U.S.A. 72 (2), 737-741 (1975)  
 JOURNAL MEDLINE  
 PUBMED 1054851  
 AUTHORS Sugimoto, K., Sugisaki, H., Okamoto, T. and Takamami, M.  
 Studies on bacteriophage fd DNA. IV. The sequence of messenger RNA  
 for the major coat protein gene  
 J. Mol. Biol. 110, 487-507 (1977)  
 JOURNAL MEDLINE  
 PUBMED 1054851  
 AUTHORS Beck, E., Sommer, R., Auerwald, E.A., Kutz, C., Zink, B., Oetelburg, S.,  
 Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takamami, M.  
 Nucleotide sequence of bacteriophage fd DNA  
 Nucleic Acids Res. 5 (12), 4495-4503 (1978)  
 JOURNAL MEDLINE  
 PUBMED 745987  
 AUTHORS Gray, C.P., Sommer, R., Polke, C., Beck, E. and Schaller, H.  
 Structure of the origin of DNA replication of bacteriophage fd  
 Proc. Natl. Acad. Sci. U.S.A. 75 (1), 50-53 (1978)

QY 1 ATGATCAAGCTGAGATCAAGCCGACGCCCGCTTACACACCCGAGCGGCTTACG 60  
 DB 1470 ATGATCAAGCTGAGATCAAGCCGACGCCCGCTTACACACCCGAGCGGCTTACG 1529  
 QY 61 CGGAGGCGAGCGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 120  
 DB 1530 CGGAGGCGAGCGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 1559  
 QY 121 TACCGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 180  
 DB 1590 TACCGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 1649  
 QY 181 TACACCGTGAAGCTGAGCGGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 240  
 DB 1650 TACACCGTGAAGCTGAGCGGCTTACACACCCGAGCGGCTTACACACCCGAGCGGCTTACG 1705  
 QY 241 CTGCGGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 264  
 DB 1710 CTGCGGCTGCTGGTGAAGATCAACCTGGAGCGAGCGGCGGCTTACACACCCGAGCGGCTTACG 1733

MEDLINE 78116039  
 PUBMED 272686  
 REFERENCE 6 (sites)  
 Horabin, J., and Webster, R.E.  
 An amino acid sequence which directs membrane insertion causes loss  
 of membrane potential.  
 J. Biol. Chem. 263 (23), 11575-11583 (1988)  
 JOURNAL MEDLINE  
 PUBMED 8829819  
 COMMENT Original source text: fd from E.coli.  
 (1) sites: internal signal sequence.  
 (2) sites: single-stranded circular DNA which codes for ten proteins  
 replicative form is duplex, filamentous.  
 (3) missing data project.

FEATURES  
 source  
 1..6408  
 /organism="Enterobacteria phage fd"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10864"  
 join(6007..6408,1..831)  
 /note="11 (x5 replication, nicking)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03303.1"  
 /db\_xref="GI:215395"  
 /translation="YIDMLVLRPISDLVOSRLSGNDLAFVDSKIATSGKNLSA  
 RVEVHIDCLTVSGLSHFESLPYHSGIAFKIYESGKNYPCVEIKASPAKVLQGH  
 NVGTGTDALGSEALLNPANSLPCLVLDLNNATIRIRATFASAPNENIACVCI  
 DILRWNSQOTRSTSRQMSRSTVNEISRHRTIYAIKHEBLQDVOQLSKPRKAM  
 TSYKEQLVLSNPDLRFASGVVFEKATIEKVLKSGPLNFDARFASDVYRQJ  
 KOLIFLWSFGFSELFKAFESGKNIVDSAVLQSKHPTTPSGKTSFKARVYF  
 GYRRLVWEGYDSVALTPWRFNSFRVYSAVLVECGIPKQQLNLSYCRVYPLVERIN  
 DFSQRPDWVNEPKIA"  
 496..831  
 /note="X (function unknown)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03304.1"  
 /db\_xref="GI:215396"  
 /translation="KNYEDSAVLEAIOSKRTTPSGTSFAKASRYEGFYRLVNE  
 GDSVALTPRNSFRVYSAVLVECGIPKQQLNLSYCRVYPLVERIN  
 VNEPLKIA"  
 843..1106  
 /note="V (single-stranded binding protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03305.1"  
 /db\_xref="GI:215397"  
 /translation="KKEIKPSQAGTTRSGVSRQCKVYSLNEOJCIVDLNBYVUL  
 VITLDEQFAYAPGLYTVHSSFKVGFQGLMIDRLVPAK"  
 1108..1209  
 /note="VII (morphogenesis)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03306.1"  
 /db\_xref="GI:215398"  
 /translation="WEQVACPDFTVCAMIQISVVLCPALGIIINAGQR"  
 1206..1304  
 /note="IX (minor coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03307.1"  
 /db\_xref="GI:215399"  
 /translation="SVLVYFASFEVLCMLRSGTTFYTRIMETSS"  
 <1301..1551  
 /note="mRNA (3' end +/- 14 bp)"  
 1301..1522  
 /note="VIII (major coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03308.1"  
 /db\_xref="GI:215400"

CDS  
 1..6408  
 /organism="Enterobacteria phage fd"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10864"  
 join(6007..6408,1..831)  
 /note="11 (x5 replication, nicking)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03303.1"  
 /db\_xref="GI:215395"  
 /translation="YIDMLVLRPISDLVOSRLSGNDLAFVDSKIATSGKNLSA  
 RVEVHIDCLTVSGLSHFESLPYHSGIAFKIYESGKNYPCVEIKASPAKVLQGH  
 NVGTGTDALGSEALLNPANSLPCLVLDLNNATIRIRATFASAPNENIACVCI  
 DILRWNSQOTRSTSRQMSRSTVNEISRHRTIYAIKHEBLQDVOQLSKPRKAM  
 TSYKEQLVLSNPDLRFASGVVFEKATIEKVLKSGPLNFDARFASDVYRQJ  
 KOLIFLWSFGFSELFKAFESGKNIVDSAVLQSKHPTTPSGKTSFKARVYF  
 GYRRLVWEGYDSVALTPWRFNSFRVYSAVLVECGIPKQQLNLSYCRVYPLVERIN  
 DFSQRPDWVNEPKIA"  
 496..831  
 /note="X (function unknown)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03304.1"  
 /db\_xref="GI:215396"  
 /translation="KNYEDSAVLEAIOSKRTTPSGTSFAKASRYEGFYRLVNE  
 GDSVALTPRNSFRVYSAVLVECGIPKQQLNLSYCRVYPLVERIN  
 VNEPLKIA"  
 843..1106  
 /note="V (single-stranded binding protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03305.1"  
 /db\_xref="GI:215397"  
 /translation="KKEIKPSQAGTTRSGVSRQCKVYSLNEOJCIVDLNBYVUL  
 VITLDEQFAYAPGLYTVHSSFKVGFQGLMIDRLVPAK"  
 1108..1209  
 /note="VII (morphogenesis)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03306.1"  
 /db\_xref="GI:215398"  
 /translation="WEQVACPDFTVCAMIQISVVLCPALGIIINAGQR"  
 1206..1304  
 /note="IX (minor coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03307.1"  
 /db\_xref="GI:215399"  
 /translation="SVLVYFASFEVLCMLRSGTTFYTRIMETSS"  
 <1301..1551  
 /note="mRNA (3' end +/- 14 bp)"  
 1301..1522  
 /note="VIII (major coat protein)"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="AA03308.1"  
 /db\_xref="GI:215400"



CDS

/protein\_id="CAC05199.1"  
/db\_xref="GI:9928951"  
/translation="MSVLVPSFASFLVKCLRSGITVFTRLMESGSYSL"  
1301..1522  
/note="gene VII1"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC05200.1"  
/db\_xref="GI:9928953"  
/translation="EKKRTGYTTVDLSQMRKEFEAFQSVAOCTTYQTQLDITAE  
LKVKNKKIFNYPARSHIARIIMAHPEFMNAKDGLIWDSYHPCYTVEHQETFRP  
SNMSEFAVDFECPFLHIYSODVACYGMLAIFPKGPIENMFVSANPMWSTSDLVW  
ANNDFEAFVPEMGKYVTQGDKWAMPILAIVHAACDPHFVGAMINELQQVCDEMGGG  
A"

-35\_signal  
-10\_signal  
misc\_feature

CDS

/note="ompA FLAG-jun-gene ::lc"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC05202.1"  
/db\_xref="GI:9928954"  
/translation="MKCAIAIAVALAGFAIVAQDYKDVDAGRIAREEVKTLXAA  
QNSBELSTANMLREQVAOLKORVMHGGAENAGCGSGSGSGSGSGSGSGSGSGS  
GSGSGSGSGSGSGDFDEKMANNAKGMTEBNEDNLLOSDAKGLDISATYDGAILE  
GFIDTSGSIANGNGATGDFPAGSNSOMAGCGNSPLMNPFROYLPSPQGVETRPPV  
FSAGKXPYSIGDCXNLNFRGFAPLIYVATRYVFTSFANILRNKES"  
2607..2669  
/note="ompA signal sequence"  
2670..2681  
/note="FLAG peptide"  
2697..2816  
/note="jun peptide"  
2832..3404  
/note="gene III C-terminal domain"  
3446..3488  
/note="fd terminator"  
3503..3841  
/note="gene VI"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC05203.1"  
/db\_xref="GI:9928955"  
/translation="MPVLLGIPLLRLRPGLVLTDFGYLLTLKKGFGXIALSLFE  
ALLIGNSTINGLSDISAOQPSDFVQGVCIILPSNALPGFVILSVKAALFFEDVQX  
KIYSYLDMGX"  
3844..4890  
/note="gene I"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC05204.1"  
/db\_xref="GI:9928956"  
/translation="MAVVPVFGKLGSGKTYSVNGCDRIKVGCKIATNIDLRCJNP  
OVNRFPKTPRVLRIPDXSIGDLAIAGNDSDYDENKNGILLDEGGRFNTRESNNKR  
EROPITDWLRLARKLGMDIFLVDDLSIDEQAASAIEHVYCRRLDPLTFPVGITL  
YS-VTSNGSLHLRHGVGVKYGDSQLSPFVERWLVTGKLINYADTKQAFASNVDGCV  
VKSIVQSYTDPRKPTIDSQRNLSTRYVEFDKSKLINSDIQKQGSITYILDCTVNCN  
SIKQNSNIIVCN"  
4868..6148  
/note="gene IV"  
/codon\_start=1  
/transl\_table=11

CDS

/note="gene IV"  
/codon\_start=1  
/transl\_table=11

BASE COUNT	1769 a	1433 c	1492 g	2357 t	4 others
Orig.N					
Query Match	55.2%	Score 145.6	DB 6	Length 7055	
Best local Similarity	72.0%	Pre. No. 14e-13			
Matches 100	Conservative	0	Mismatches 74	Indels 0	Gaps 0
Db	1	ACGATCAGAGTGGAGATACAGCCAGCCAGCCAGCTTACACCTCGCAGCGGCTTACG	63		
Cy	843	ATGATTAAAGTTGAAGTAAACCGCTCGACGCCAATTACAGCCGCTTGGTATTTCT	922		
Db	61	CGCAGGGGAGGCCCTACAGCTGAAAGAGCAGCTGGCAGGTGACACTCGGCAGACAG	120		
Cy	903	CGTCAGGCGAGCCTTATTCAGCTAAAGACAGCCTTGTACGTTGATTGGTAA	962		
Db	121	TACCCGCTGCTGTGTAAGATACCCCTGAGAGAGGCGACGCCCGCTACGCGCGGCTG	180		
Cy	963	TATCGGAGCTGTGAAGATATCTCGAGAGAGGTAGCAGAGCGATGCGGCTGGTCTG	1022		
Db	161	TACACCGTGCACTTGAGAGCTTAAAGTGGCCATTTGGCAGCGCTATATGATCAGCGC	240		
Cy	1023	TACACCGTGCACTGTCTCTGCTTAAAGTTGGTCAAGCTGGCTTTTATGATTAACGCT	1082		
Db	241	CTGCGCGCTGTCGCCGCCAATTA	264		
Cy	1083	CTGCGCGCTGTTCCGCGTAATTA	1106		
RESULT 8					
BD131874					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
ORIGIN					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					



```
CDS
496..811
/codon_start=1
/trans_table=11
/product="replication protein px"
/protein_id="AA073896.1"
/db_xref="GI:21667081"
/translaction="MNIDDSAVLDAIQSKHFTIPDSKTSAPKASRFGPRVJNE
CNSVLTMPRSFMRVSAIVECGIPKQCLNLTGNVPLVAFINVDSSQPFCH
MNPVLIKIA"
843..1106
/codon_start=1
/transl_table=11
/product="cyclicolic single-stranded DNA binding protein
pv"
/protein_id="AA073897.1"
/db_xref="GI:21667392"
/translaction="K:KVEIKPSQAQFTIRSVSGQKPEVINEQCCVYAGNEPV
VKITLDEGPAVAPGLYVTHSSFKVQCFQS.MIDRLRVPK"
1108..1209
/codon_start=1
/transl_table=11
/product="minor coat protein pvt"
/protein_id="AA073898.1"
/db_xref="GI:21667083"
/translaction="MEQVADFDIYQAM:QISVYLCFALGIIAGGCR"
1206..1304
/codon_start=1
/transl_table=11
/product="minor coat protein ptx"
/protein_id="AA073899.1"
/db_xref="GI:21667084"
/translaction="MSVLYVSFASVFLGWCLSGCTFTRLMETSS"
1301..1522
/codon_start=1
/transl_table=11
/product="major coat protein pvti precursor"
/protein_id="AA073900.1"
/db_xref="GI:21667085"
/translaction="MKKSLIKASVAVATLVPLSPAEEDPAAVAPDS:QASATEY
IGYAMVAVIVGATIGIKLFFKFTGKAS"
1579..2853
/codon_start=1
/transl_table=11
/product="adsorption protein pili precursor"
/protein_id="AA073901.1"
/db_xref="GI:21667086"
/translaction="MKKILFALPLVFPYSHSAETVESCLAKPHTESTFWKDKT
LDRVAYEGCLMAYGVCTGDETCGTVWVIGLAIPENGGSGEGSGGSEGG
GRTKPEYGDPIPGYTYINLDDGTYPGTGWPANPNSLESDPLNFMKNRFR
NRGCACTVYGTCTGCTDPYKTYVYTPVSSKAMTAYNKKFDRDARSGNTEFFV
CEIQQSSDLPPPVNAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
DYENANANAGATENADENALQSLAKGLDSVATVGAATGGF:GVYSGLMAGAT
GDFAGNSQMAQVGDGNSPLMNFQVYPSLPOSVECRFVYFCAGKPYEFSDCCX:
NLFRGVAFLLVATEWYVFSTANILRNKES"
2856..3194
/codon_start=1
/transl_table=11
/product="minor coat protein pvt"
/protein_id="AA073902.1"
/db_xref="GI:21667087"
/translaction="MPELGIPLLRFLGLVTFGYLTLFKKGFCAIAISLF
ALIGNSLIVGYLDSISQLPSDFVQVQLILPSMALPCFVILSVKAAITFVQK
KIVSLDMCK"
3197..4243
/codon_start=1
/transl_table=11
/product="morphogenetic protein pi"
/protein_id="AA073903.1"
/db_xref="GI:21667088"
/translaction="MAVYFVTKLGGKTVYSGKIDKIVAGCKATNLDKQMLP
QVGRFACFPYLRIPKPSISDLA:GRGNDSDYKNGKGLVLDGCTFNTSRNDK
ERQPTDFLHARLGDIIITLVQDD:SVDKCARSLAAHVAYCRLLDTLPFGVL
YSLVTGSKMPLPKLHVGVVYVYQDLSPTVERWLVTGKXLYNAYDTKQAFSSNVDG
```

```
CDS
VSVLTPEYLSHGRYKPEKLNQQRKLTXYIKKPSNVLCTA:GFASAPFVSY:TOPKPE
VKKVSTQYIDFKFTIDSSQRLNLSYRIVFQSGKGLKNSDLQKQSGITVLDLCTV
SIRKQNSSEIVKCN"
4221..5501
/codon_start=1
/transl_table=11
/product="morphogenetic protein lv"
/protein_id="AA073904.1"
/db_xref="GI:21667089"
/translaction="VKLNATVNFVEMVSSSPQVLEMMNSF:K:R:VYVYSKQTS
SVYSPDVKGTIVVSSCEVPEINFEISV:RANPEMWGSEITISIKVYNSQDYI
DELPSSD:GEYEDNAPSQGFPCNCTVOTFKINVRACDLIRVCLFVNSKTS
SNVLSYDGSN:LVASAPKIDLM:PCFLSTVLELPPDOL:ELL:FEVQGDALPESA
ASQSGTVAGVYDTR:SVLS:SSAGSGSG:ENGQIG:SVRAIKYNSKSLSVRL
TISQKGSISVQGVNPFITGRITGSAVANNPFCVIRQVQNGISNCGVPMAGNIV
LITSKADSLSSVQASVITNQRSLA:TVLRDQCTLLGLTYNNNTSCSGVPEL
SKIP:GLFSSRSISNESTLYLVKAT:VVAL"
5878..6537
/codon_start=1
/transl_table=11
/product="chaperoninco: acety: transferase"
/protein_id="AA073894.1"
/db_xref="GI:21667090"
/translaction="MEKITGYTVVJISQMKRKEHEAFQVQCTVQVQLDITAF
:KTVKKNKREYPAE:HILAR:KNARPEFRAMKQGEVLEINOS:TRCVYFHQCTET
SLMSEYEDDFRQPLHIVSOVAVAGEN:AVFPKGF:ENVFVSANPWSFPLANV
AMDNDFAPVFTMGVYTGQKVLMLPLAIVHVAVCTGFHVGMMALRLQVCDERQGG
A"
BASE COUNT 1962 a 1585 c 1641 g 2587 t
ORIGIN
Query Match 55.2%; Score 145.6; DB:2; Length 7775;
Best Local Similarity 72.0%; Pred. No. 1.3e-13;
Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY : ATGATTAAGTGAAGTCAAGCCACGCCAGGCCGCTTACACACCCGCGAGCGCTGAGC 60
DB 843 ATGATTAAGTGAAGTCAAGCCGCTTACACACCCGCTTACACACCCGCTTACAC 902
QY 61 CGCAGGAGGAGCGCTTACACACCGAGCGCTTACACACCGAGCTTACACACCGAG 120
DB 903 CGCAGGAGGAGCGCTTACACACCGAGCGCTTACACACCGAGCTTACACACCG 962
QY 121 TACCCGCTGCTGTAAGATCAACCTGAGAGGAGCGCGAGCGCGCTTACACCGCTG 180
DB 963 TACCCGCTGCTGTAAGATCAACCTGAGAGGAGCGCGAGCGCGCTTACACCGCTG 1022
QY 181 TACACGCTGACCTGAGAGCGCTTCAAGGCTGCGCGAGCTGAGAGCTGATGAGAGCG 240
DB 1023 TACACGCTGACCTGAGAGCGCTTCAAGGCTGCGCGAGCTGAGAGCTGATGAGAGCGT 1082
QY 241 CTGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 284
DB 1083 CTGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
RESULT: 0
AX006302 7793 bp DNA linear PAT 24-AUG-2000
LOCUS AX006302
DEFINITION Sequence 3 from Patent WO9906587.
ACCESSION AX006302
VERSION AX006302.1 GI:1992893.
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1. Ge... and Jia... V.
AUTHORS Novel method and phase for the identification of nucleic acid
CITILE sequences encoding members of a multimeric (poly)peptide complex
JOURNAL Patent: WO 9906587-A 3 11-FEB-1999;
FEATURES
GE LIMITING (DE); ILAG VIC (DE)
Location/Qualifiers
```

```
source
1..7783
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/feature="phage vector (circular)"
<1..831
/feature="C-terminus gene 11"
/codon_start=1
/transl_table=1
/protein_id="CAC05183.1"
/db_xref="GI:928933"
/translation="NATTSRIDATRSAPAPNENAKGVICHURVNSNGOTSTSSCN
WESTVWNTSRRTUVAHLKHVELQDQCCJSSKSAKTSYCKQLVSNPDLIE
PAGSLVPEARLERTYLSFGJLNLDAIRFSGJNPOCKDIFDLNPFSSJLKA
PEGSNNIYDSSAVJDAIGSHFTTTPSGKTSFNKASRYGFRJLVNCGYSVALTM
PRNSFWRYVALVEGCI:PKSQLMNLSCNNVPLVRFINVDPSORPDMVNEPVLKIA"
496..831
/feature="gene X"
/codon_start=1
/transl_table=1
/protein_id="CAC05184.1"
/db_xref="GI:928933"
/translation="MIYVDSAVLDAIQSHFTTTPSGKTSFAKASRYGFRJLVN
GYDSVALTMPRNSFWRYVALVEGCI:PKSQLMNLSCNNVPLVRFINVDPSORPDM
VNEPVLKIA"
843..1106
/feature="gene V"
/codon_start=1
/transl_table=1
/protein_id="CAC05185.1"
/db_xref="GI:928933"
/translation="MIKVEIKPSQAOFTTSGVRCGKPYSLNEQLCYVDLNEVYL
VKITLDEGAPYARGLYVHLSSFKVQGFPSLMLIDRLRLYPAK"
1198..1203
/feature="gene VII"
/codon_start=1
/transl_table=1
/protein_id="CAC05186.1"
/db_xref="GI:928933"
/translation="MEGVADFDITVQAMIQ:SVVLCFALG:TAGGQR"
1206..1316
/feature="gene IX"
/codon_start=1
/transl_table=1
/protein_id="CAC05187.1"
/db_xref="GI:928933"
/translation="MSVLVSPASFY:GMCLRSGLTYFTPLMETSSCVSL"
1301..1522
/feature="gene VII"
/codon_start=1
/transl_table=1
/protein_id="CAC05188.1"
/db_xref="GI:928933"
/translation="VRLKAVINPVFXVSSSEFAGVLENNSPJRFPTWVSKQJGS
SIVSPVQKGVTVSSDVCKPEVLRNFSLVLRANFEWGSIPSTCYVNSQCYI
DELPSRICEYDNSAPSGGFVQNDWOTFKIKNNAKQITRYELFVNSCTKS
SNHSVNGMLLVASAPKCTLNLDPFLSNOLPTKQOILIEZIFPVCGDALDEPFA
AGSRGCVAGCVTDLTSLVLSAGSGFSGFNQVGLSVRAKCNBSKXITSVRTIL
SLSQKGISVGGVVPITGRVIGESANVNPQVTERQNVG:SMSVFFVAAAGCNIV
LDTITSKADS:SSSTOASDVITNORSIAITVNLDDGTLILGGITDYKXNSJDSGVPF
SKIPICGLFSSRSRDSNEESTLVYLKATIVRAJL"
1643..2302
/feature="cat resistance gene"
/codon_start=1
/transl_table=1
/protein_id="CAC05189.1"
/db_xref="GI:928933"
/translation="MEKKITCYTVVD:SCWHRKEHFEAFQSVACQVYNOTVQLD:CAF
IKTVKAKHKFPYPAFHILARLMAHFEFFMAKDDGLVWDSVHPCYVFFHQETTF
SLSWSEHDDRFQPHIVYSQDVACVGENLAYFPKGFLENMFVSVANPWSFTSFDLNV
ANMDFAPVFTMGKYVTOGKVLNPLAIOVHAHVCDFGVHGMELNELQYCGEMQCG
A"
2656..2669
/feature="10-signal"
2679..2683
/feature="1ac operator"
2769..4139
/feature="ompA-FLAG-scfv (anti-HAG)-gene 11ss"
BASE COUNT 1920 a 1649 c 1667 g 2543 t 4 others

/codon_start=1
/transl_table=1
/protein_id="CAC05190.1"
/db_xref="GI:928933"
/translation="VKKRTATLAVLAGFAVACADYCTVWTSRSLITPAGSKYT
VSCITSSQSLNPSCKQNYTWQKFGQPPKLTWAS:RSGVDFRTGSSGCTDFT
LITTSVQAEDELAVYCONVSNPCTFGNSTRLELFAJGGGSGKSGGSGGSG
GGSGGGGSEVLVYESGGDLYNPGGSLKLSQASGSESSGSMVRCPTPDXRLEWVA
TISNGGTYVYVPSVKGFRFTSPDNKAKTLIYOMSLKSGDSAVYVCAHREYDEGFA
AWGQGTLVTVSASGEFFSAGAGGSGSGGDFPEKXANAKCAVETNADKALQSDA
KGLLSVATVYGAAIDGFTGDVSGLAGNGATGFFRGSNSQVACVGDGPNSLMNP
CY:PSLPOSVECFPPVFGAKKPEFSIDCKRINLRFQVAFILLYATRWYVISTFANK
LRKRES"
2769..2831
/feature="ompA signal sequence"
2832..2843
/feature="FLAG peptide"
2844..3641
/feature="scfv (anti-HAG)"
3666..4139
/feature="gene 11ss"
4178..4220
/feature="fd terminator"
4231..4569
/feature="gene V"
/codon_start=1
/transl_table=1
/protein_id="CAC05191.1"
/db_xref="GI:928933"
/translation="MPVLLGIPLLPFLGFLVTV:FGVLLTF:KSGFGKIAIA:SLFL
ALII:GNSILVGLSDISAKLPDEFVQGVQLLPSPALPFPVILSVKAAIFFDVQK
KIVSYLDMQK"
4572..5638
/feature="gene I"
/codon_start=1
/transl_table=1
/protein_id="CAC05192.1"
/db_xref="GI:928933"
/translation="VAVVFTGVLGSGKTVSGK:CDXIVAGCKIATYDLRLOLP
OVGRFAKTPVVRIPDPSISDLA:GRNDSDYEFKKQ:LVNDEGGTFMNPSSMKK
BROPILIDFLHAKKLMDI:FLYQD:SLYDKOKRSLAHVYCRLEDLITLPRVETL
YSKTSKMPLEKLVGVVKIGDSJLSPVVERLVYKLVNAYDTRKASPSISBGV
YSILTPYLSHGRYFKPLNIGQMKLTKIYAKRSRYLCAIGFASFTYSITQPPS
VKVNSCTVDFEKFITIDSSORLNL:SYRVYFKSGKGLINSDELQKQVSIIV:DLCTV
S:KKXGNSBIYKCN"
5556..6876
/feature="gene IV"
/codon_start=1
/transl_table=1
/protein_id="CAC05193.1"
/db_xref="GI:928933"
/translation="VRLKAVINPVFXVSSSEFAGVLENNSPJRFPTWVSKQJGS
SIVSPVQKGVTVSSDVCKPEVLRNFSLVLRANFEWGSIPSTCYVNSQCYI
DELPSRICEYDNSAPSGGFVQNDWOTFKIKNNAKQITRYELFVNSCTKS
SNHSVNGMLLVASAPKCTLNLDPFLSNOLPTKQOILIEZIFPVCGDALDEPFA
AGSRGCVAGCVTDLTSLVLSAGSGFSGFNQVGLSVRAKCNBSKXITSVRTIL
SLSQKGISVGGVVPITGRVIGESANVNPQVTERQNVG:SMSVFFVAAAGCNIV
LDTITSKADS:SSSTOASDVITNORSIAITVNLDDGTLILGGITDYKXNSJDSGVPF
SKIPICGLFSSRSRDSNEESTLVYLKATIVRAJL"
6874..6952
/feature="packaging signal"
6992..7117
/feature="fd ori"
7382..73793
/feature="N-terminus gene 11"
/codon_start=1
/transl_table=1
/protein_id="CAC05194.1"
/db_xref="GI:928933"
/translation="KIDMLVRLPFLDSLVCSR:LSGNDLIARVYDLCK:ITLSCANLSA
RVEYHIDGDLTVSGLSHPPELPHVSG:AFKTYGSKNFVPCVEIKASPAKVLQGH
NVFGITDLACSEALLINPANSIPCLVYDLDV"
BASE COUNT 1920 a 1649 c 1667 g 2543 t 4 others
```





TITLE Direct Submission  
JOURNAL Submitted (18-XAR-2001) Cell Research and Immunology, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel  
FEATURES  
SOURCE 1. 8233  
Location/Qualifiers  
/organism="Peptide display vector fct1"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:160651"  
/note="type 88 filamentous phage display vector derived from bacteriophage fd"

BASE COUNT 2026 a 1651 c 1780 g 2776 t  
ORIGIN

Query Match 55.2%; Score 145.6; DB 12; Length 8233;  
Best Local Similarity 72.0%; Pred. No. 1.3e-13;  
Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

1 ATGATCAAGTGGAGATCAAGCCGACGACGCGGACGCTGACACCGGACGCGGACG 60  
DB 1470 ATGATCAAGTGGAGATCAAGCCGACGACGCGGACGCTGACACCGGACGCGGACG 1529  
QY 61 CGCCAGGCGACGCGCTTACAGCGCTGACAGAGAGCGCTGCTACGCTGAGACCTGCGGACG 120  
DB 1530 CGTCAAGCGACGCGCTTATCTACTGATGAGAGAGAGCTTCTTACGCTGATTTGCGATGAA 1589  
QY 121 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 180  
DB 1590 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 1649  
QY 181 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 240  
DB 1650 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 1709  
QY 241 CTGGCGCTGCTGCG 264  
DB 1710 CTGGCGCTGCTGCG 1733

RESULT 13  
AF218733 8879 bp DNA circular SYN 25-JUN-2002  
LOCUS Filamentous phage display vector fMCS1, complete sequence.  
DEFINITION AF218733  
ACCESSION AF218733  
VERSION AF218733.1 GI:8698958  
KEYWORDS  
SOURCE  
ORGANISM Filamentous phage display vector fMCS1  
REFERENCE  
AUTHORS Smith G.P. and Fernandez, A.-M.  
TITLE Direct Submission  
JOURNAL Submitted (22-DEC-1999) Division of Biological Sciences, University of Missouri, Tucker Hall, Columbia, MO 65211-7403, USA  
COMMENT fMCS1 is a filamentous phage cloning vector derived from fd-tet by replacing 329 non-essential base-pairs with a 25-bp multiple cloning site. fMCS1 is similar in all important respects to its fd-tet parent. Minus strand synthesis is slow, so double-stranded RF copy number is low (about 1 in growing cells; 100 in stationary-phase cells). Yield of physical particles is about 5e11 virions/ml. Infection with fMCS1 transduces tetracycline resistance into the host cell, and the vector can be propagated like a plasmid by growing infected cells in medium containing tetracycline. Plaques are very small. Foreign inserts can be spliced into the unique restriction sites in the multiple cloning site, and are generally stable.

## FEATURES

source  
Location/Qualifiers  
1. 8879  
/organism="Filamentous phage display vector fMCS1"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:127589"  
/lab\_host="Escherichia coli"  
/locus  
join(1..5646,5672..8879)

CDS  
/organism="Filamentous phage cloning vector fd-tet"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:127589"  
/note="presented in Genbank Accession Number AF217317"  
4221..5501  
/codon\_start=1  
/transl\_table=11  
/product="protein 1"  
/protein\_id="AA078531.1"  
/db\_xref="GI:8698959"  
/translation="MKLVNINVEFLIMFSSSSFAQVLEVNNSPLRDETVYSKQTE  
SVTSPPDKQTVIVSSPKPKENLNFISVRRANFPVGSIBELIKVYNADQY  
DELPSDIDEDFNAPSGGFVPCQDNVTCFKINVRAKHISAVEFKVNSKS  
SVLSDVGNLAVSAFKQIDNLPQFSLVLPDQILLEGILFVQGMALPFSFA  
AGQQRGTAVGVTDLTGVSSAGSGECPFGQDVGILSVRLKTNKSHKILVPRIL  
TSGQXGSLISVGVNPFITGRVGSANVNPQVVERQVGLISVFPVMAAGNIV  
LITSKASLSSTQADVTINQRIATVNRDQITLLGSLTYKVTISQSGVRF  
SKPLGLGIPSSRSDNESTIVYKNTIVRAL"

misc\_feature  
BASE COUNT 2409 a 1822 c 1803 g 2843 t  
ORIGIN  
Query Match 55.2%; Score 145.6; DB 12; Length 8879;  
Best Local Similarity 72.0%; Pred. No. 1.3e-13;  
Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

1 ATGATCAAGTGGAGATCAAGCCGACGACGCGGACGCTGACACCGGACGCGGACG 60  
DB 843 ATGATCAAGTGGAGATCAAGCCGACGACGCGGACGCTGACACCGGACGCGGACG 903  
QY 61 CGCCAGGCGACGCGCTTACAGCGCTGACAGAGAGCGCTGCTACGCTGAGACCTGCGGACG 120  
DB 903 CGTCAAGCGACGCGCTTATCTACTGATGAGAGAGAGCTTCTTACGCTGATTTGCGATGAA 962  
QY 121 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 180  
DB 963 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 1022  
QY 181 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 240  
DB 1023 TACCGCGCTGCTGAGAGATCAAGCCGACGAGGCGCGCGCGCTACCGCGCGCGCTG 1092  
QY 241 CTGGCGCTGCTGCG 264  
DB 1083 CTGGCGCTGCTGCG 1106

RESULT 14  
AF217317 9183 bp DNA circular SYN 19-NOV-2002  
LOCUS Filamentous phage cloning vector fd-tet, complete sequence.  
DEFINITION AF217317  
ACCESSION AF217317  
VERSION AF217317.1 GI:5649178  
KEYWORDS  
SOURCE  
ORGANISM Filamentous phage cloning vector fd-tet  
REFERENCE  
AUTHORS Beck E., Scherer R., Auerwald E.A., Kurz G., Zink B., Osterburg G., Schaller H., Sigmund K., Sasaki H., Okamoto C. and Takamori M.  
TITLE Nucleotide sequence of bacteriophage fd DNA  
JOURNAL Nucleic Acids Res. 5 (12), 4495-4503 (1978)  
MEDLINE 75136480  
PUBMED 745987

REFERENCE  
1 (bases 1 to 5645; 8421 to 9183)  
Beck E., Scherer R., Auerwald E.A., Kurz G., Zink B., Osterburg G., Schaller H., Sigmund K., Sasaki H., Okamoto C. and Takamori M.  
TITLE A new filamentous phage cloning vector: fd-tet  
JOURNAL Gene 9 (1-2), 127-140 (1980)  
MEDLINE 80202561  
PUBMED 6247241  
REFERENCE  
3 (bases 5500 to 5577; 5624 to 5645; 8421 to 8532; 8541 to 9680)

AUTHORS  
Zinder, N.E. and Horvath, K.  
TITLE  
Multiregulatory element of filamentous bacteriophages  
JOURNAL  
Microbiol. Rev. 49 (2), 101-106 (1985)  
MEDLINE  
85240391  
PUBMED  
289671  
REFERENCE  
4 (bases 1 to 9183)  
AUTHORS  
Smith, G.P.  
TITLE  
Filamentous phage assembly: morphogenetically defective mutants  
JOURNAL  
Virology 167 (1), 156-165 (1988)  
MEDLINE  
89045639  
PUBMED  
3183194  
REFERENCE  
5 (bases 1 to 9183)  
AUTHORS  
Smith, G.P.  
TITLE  
Filamentous phage vector fd-tet  
JOURNAL  
Unpublished  
REFERENCE  
6 (bases 1 to 9183)  
AUTHORS  
Smith, G.P.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (16-DEC-1993) Division of Biological Sciences, University  
of Missouri, Tucker Hall, Columbia, MO 65211-7400, USA  
COMMENT  
Filamentous phage cloning vector constructed by splicing the  
tetraacycline resistance determinant of T10 into the minus-strand  
origin of bacteriophage fd. Minus strand synthesis is slow, so  
double-stranded RF copy number is low (about 1 in growing cells;  
100 in stationary-phase cells). Yield of physical particles is  
about 5e11 virions/ml. Transduce tetraacycline resistance into  
infected host cell. Make very small plaques. Foreign inserts can  
be spliced into unique HindIII site, and are generally stable.  
FEATURES  
source  
1..9183  
Location/Qualifiers  
1..9183  
/organism="Filamentous phage cloning vector fd-tet"  
/mol\_type="other DNA"  
/db\_xref="taxon:127516"  
/lab\_host="Escherichia coli"  
/foci=5  
1..3645  
/organism="Enterobacteriophage fd"  
/mol\_type="other DNA"  
/db\_xref="taxon:13864"  
8421..9183  
/organism="Enterobacteriophage fd"  
/mol\_type="other DNA"  
/db\_xref="taxon:10864"  
join(8782..9183,1..831)  
/codon\_start=1  
/transl\_table=1  
/product="replication protein pII"  
/protein\_id="AA75600.1"  
/db\_xref="GI:8489183"  
/translation="MIDMLVLRPFDLSVGRSLGNDLIAFVDSKIATLSGNLSA  
RTVEYHDDGLTVSGLSHPESLPYVSGAFKIEGSRNFYCVIEIKASPAKVIQGH  
NVRGTDILAGSEALHNFANSJPCLYDLDVAATTISRIDATPSARAPNEIACV  
DHLRNVNGQTKSTKSONMESITVNETSRHRTLVAVLGHVLELOHQIQLSSEPAKQ  
TSYQEKQLKVLNSPDLLEFASGLVREARLIEIRYKSLPLNLFDAIRFADYRQG  
KDIPLMSRSPSELKAFKPGDSNIVYDSAVDAIOSGRTITPSGKTSFAKARIT  
GPRRLVNEGYSVALTMRPNRFRYVSLVVEGGLPKQGLMNLSTOMNVPVLRFINV  
DESSGRDWTNEVLEKIA"  
496..831  
/codon\_start=1  
/transl\_table=1  
/product="replication protein pX"  
/protein\_id="AA75601.1"  
/db\_xref="GI:8489183"  
/translation="NNIYDSSAVLDAIQKRPFTTPSGKTSFAKARIFGFRRLVNE  
GYDSVALTMRPNRFRYVSLVVEGGLPKQGLMNLSTOMNVPVLRFINVDESSGRDWT  
NEVLEKIA"  
843..1106  
/codon\_start=1  
/transl\_table=1  
/product="cytosolic single-stranded DNA binding protein  
pV"  
/protein\_id="AA75603.1"  
CDS  
/db\_xref="GI:8489184"  
/translation="MIRKEIVKSGAQFTTRSGVSGGKFRYSNEELCYVDQNEEYPLV  
KRIIPGGQPAVAPGLYIVRSLSPKVGQFGSLMIDRLVLPK  
1108..1209  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pV.1"  
/protein\_id="AA75604.1"  
/db\_xref="GI:8489185"  
/translation="VEQVALFDTIVQAMQLISVLCFALGIIAGQR"  
1206..1304  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pX"  
/protein\_id="AA75605.1"  
/db\_xref="GI:8489186"  
/translation="SVSLVSSAFVSGWLRSGITFTRLMETSS"  
1301..1522  
/codon\_start=1  
/transl\_table=1  
/product="major coat protein pV.1 precursor"  
/protein\_id="AA75598.1"  
/db\_xref="GI:8489179"  
/translation="MKKSLVLSAVAVATVPELSSFAAGGCFARAAFDSCQASATY  
GVAMWVIVGATGKLFKFKTSKAS"  
1301..1369  
/codon\_start=1  
/transl\_table=1  
/product="description protein pII precursor"  
/protein\_id="AA75609.1"  
/db\_xref="GI:8489190"  
/translation="MKKLFALPLVPEFSSHAETVESCTAKPFTENSTVWKKQKT  
LDRVAVEGGLNNAQVAVCTGDETCQKVPGLALPENGSGSGSGSGSGSGSG  
SGTRPEVGDTEIPGTIYINPLEDTPTPTPEMPAKNPSLESGPLNTPKQNNRFR  
NRQGLCVITG-VTQSTDPKTYVQYTPVSSKAMPDAVANGKFRQCAEHSQNECPFV  
CEYQGGSDLPQPPVNAAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG  
DYKXANNAKGMATENADENALOSDAKGLDSVALTYGALIGPGLVSGALNMGAT  
GPAFGSSQMAQVCDGNSFLKNNRQYLPSPQSVCEHPVYQAKPYEPIIDCDK  
NLRGVFAFLIVATRTMYVSTFAKLRKES"  
1579..1632  
2856..3394  
/codon\_start=1  
/transl\_table=1  
/product="minor coat protein pV.1"  
/protein\_id="AA75606.1"  
/db\_xref="GI:8489187"  
/translation="VSVLGLFLRFLQHLVTPGVLLFLKSGFKMIAIACFL  
ALIGNSLIVGLSDISAKLPSPFQGVGLLPBALPFIIVLSKAAFIIPVKQ  
KIVSYLWDR"  
3197..4243  
/codon\_start=1  
/transl\_table=1  
/product="morphogenetic protein pI"  
/protein\_id="AA75593.1"  
/db\_xref="GI:8489180"  
/translation="SAVIVTQKLSGSKTVSVGKQCKIVAGCKIAINLDIRLONLP  
QVRFAKTPRVLRIIPKPEISDLUALISGRNDSYDNNKGLVLDSCGWFNTRSWNDK  
ERQPIDWFTHARKLGWDILFLVODISIVDQAFSALAEHVVCPRLRITLPEVGL  
YSLVGSKMPLPKJGVVAVVQDSQSLPVEKRLVYAKDYADKQAFSSNYDSGV  
YSVLYLSHGRVFKPLNIGQKMLTKILKKFSVLGIALGSAFATPYVYTKRPE  
VKKVVSQYDPPKFTITDSQRANLSRVAVKSKKLLNSDILQKQSYITIDICTV  
SIRKNSNRIKCN"  
4221..5502  
/codon\_start=1  
/transl\_table=1  
/product="morphogenetic protein IV"  
/protein\_id="AA75602.1"  
/db\_xref="GI:8489183"  
/translation="KKLVNINPFTSRSSSSPQVLEMANSGRLRFTVWYKSKTGS  
SVTSPDVAGTIVYSSDEVKPENLKNTFISVLRANFPVWGSIPSIIDKYNNSQDYI  
DELPSDIDQEDNSAPSGGFPPQNNKVTCTFKINVPKADLIVEVEFVNSNRSKS  
SNVLSYDGSNLLVSAFKDILNLDPQISTVYLPDQILTEGLPEVOGGDALPSFA

## misc\_feature

/note="morphogenetic protein signal sequence: promotes packaging of circular single plus strands into progeny virions; unclassified site"  
 join(5624..5645,8421..8532)  
 /citation={3}

## rep\_origin

/note="minus-strand origin; disrupted by Tn10-derived insert; functionally inactive; minus strand initiated randomly by endogenous bacterial Okazaki fragment priming apparatus"  
 /citation={3}

## repeat\_region

5646..8420  
 /citation={4}

## misc\_feature

5975..5983  
 /transposon="n10"  
 /note="HindIII cloning site"  
 complement(6239..7444)

## CDS

/product="Tn10 tetracycline resistance protein TetA"  
 /protein\_id="AF75607.1"  
 /db\_xref="GI:8489188"  
 /translation="MNSSTKALVCTLDANGIGTPVLPPLRFPAISDIAHFG  
 VLALYAKVTPAPLKSMSDRGRPPVLLSLGASDYLAFSEALMLYIGR.  
 LSGITGAGAAVADITTSASORVWFQWLASFGGLGADPITGGFAGEISPHSP  
 FIALNLINATVFLVVMFMRITKTRNDCTEGVETGTSVYTLFRTMPLILYF  
 SAOLIGIPATVWLFTEHFRGMSNMWGSLSAGLCHLSVPAFVAGTARKMEKT  
 AVLDGIADSSAFALAFISEGMVLPVLLLAGGIALPALAGVMSITKSHCOGAL

## Query Match

55.2% Score 145.6; DB 12; Length 9183;

## Best Local

Similarity 72.0%; Pred. No. 1,35e-13;

## Matches

190; Conservative 0; Mismatches 74; Indels 0; Gaps 0

QY 1 ATGATCAAGTGGAGATCAAGCCAGGCGCCAGCTTCAACACCCGAGCGGCTGAGC 60  
 DB 843 ATGATTAAGTTGAATTAACCGCTTCAACCGCAATTACTACCGCTTGGTCTTCT 902  
 QY 61 CGCAGGCGCAAGCCCTTACAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 DB 933 CGTCAGGCGCAAGCCCTTACAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
 QY 121 TACCCGCTGCTGGTGAAGATCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 DB 963 TATCCGCTGCTTGCAGATTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022  
 QY 181 TACACCGTGAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 DB 1023 TACACCGTGAACCTGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1082  
 QY 241 CTGCGCTGAGCGCGCCAGTAA 264  
 DB 1083 CTGCGCTGCTTCCGCTAAGTAA 1106

## COMMENT

Type-8 filamentous phage display vector constructed by engineering three single-base-pair substitutions into Filamentous phage cloning vector (0-let presented in Genbank Accession Number AF217317). 0-let is very similar in all important characteristics to its filamentous parent. It is intended as a phage-display vector. Thus the 9 bp 'stuffer' between the PstI and BamHI cloning sites can be replaced with a short degenerate coding sequence to create a library of phage displaying different random peptides at the N-terminus of a 3500 copies of the major coat protein pVIII.

## FEATURES

## source

1..9183  
 /organism="Filamentous phage display vector 0-let"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:127588"  
 /db\_host="Escherichia coli"

## gene

1301..1522  
 /gene="VIII"  
 1301..1522  
 /gene="VIII"

## CDS

/codon\_start=1  
 /translation="MKSLVAKSVAVATVPMLSFAAGSEFPAKAFDS-CASNTEY  
 IGYAMWVWVVGATIGITLFFKFTSKS"  
 1301..1369  
 /gene="VIII"  
 1370..1519  
 /gene="VIII"  
 /product="major coat protein pVIII mutant precursor"  
 /protein\_id="AF78532.1"  
 /db\_xref="GI:8698961"  
 /translation="MKSLVAKSVAVATVPMLSFAAGSEFPAKAFDS-CASNTEY  
 IGYAMWVWVVGATIGITLFFKFTSKS"

## sig\_peptide

1301..1369  
 /gene="VIII"

## mat\_peptide

1370..1519  
 /gene="VIII"  
 /product="major coat protein pVIII mutant"

## misc\_feature

1368..1373  
 /note="PstI site"

## variation

1372  
 /note="creates PstI site"

## misc\_feature

1381..1386  
 /replace="G"  
 /note="BamHI site"

## variation

1381  
 /note="creates BamHI site; results in aspartic acid to glutamic acid substitution"

## variation

2220  
 /replace="C"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

## variation

2220  
 /replace="G"

## variation

2220  
 /note="removes BamHI site"

Fri Nov 7 15:26:53 2003

us-09-622-500b-3\_1.rge

Page 17

Search completed: November 7, 2003, 12:52:43  
Job time : 1511 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic acid search, using sw model

Run on: November 7, 2003, 12:27:20 : Search time 1315 Seconds

(without alignments)  
4879.378 Million cell updates/sec

Title: US-09-622-500B-3

Perfect score: 264  
Sequence: 1 atgacaaaggtgagatcaaa.....gctcgtgcccgcacagaa 264

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapex: 1.0

Searched: 2278.392 seqs, 1215223856 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hnc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hnc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estom.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pla.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_prc.\*  
25: em\_gss\_rcd.\*  
26: em\_gss\_pbg.\*  
27: em\_gss\_vrt.\*  
28: gb\_gss1.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARYS

Result No.	Score	Query Match Length DB	Description
1	51.2	19.4	756 13 B0449043 603765817
2	50.4	19.1	493 9 AW269906 AW269906 xv37d01.x
3	50.4	19.1	493 9 A1971889 A1971889 wv29902.x
4	50.4	19.1	518 9 A1056863 A1056863 oz03ec01.x

7

5	49.4	18.7	427 10 B011441 7093653.x
6	49.4	18.7	433 14 C0828249
7	49.4	18.7	545 13 B0080351 946149508
8	49.2	18.6	454 10 B6313865 4522063.C
9	49.2	18.6	585 10 B6436680 4522063.C
10	49.2	18.6	596 12 B2251883
11	49.2	18.6	947 13 B1388416
12	49.2	18.6	945 23 B2566686
13	48.4	18.3	974 13 B0412619
14	48.4	18.2	984 29 B2551647
15	47.8	18.1	407 14 C0833570
16	47.8	18.1	443 14 C0885926
17	47.8	18.1	467 13 B0775275
18	47.8	18.1	468 13 B0494339
19	47.8	18.1	471 13 B0278825
20	47.8	18.1	479 13 B0037392
21	47.8	18.1	483 10 B6575293
22	47.8	18.1	487 13 B0079771
23	47.8	18.1	492 13 B0499994
24	47.8	18.1	493 12 B1388391
25	47.8	18.1	493 13 B0075515
26	47.8	18.1	501 12 B1675302
27	47.8	18.1	504 13 B0615678
28	47.8	18.1	511 13 B0778982
29	47.8	18.1	515 12 B1478947
30	47.8	18.1	517 13 B0779374
31	47.8	18.1	519 13 B0487062
32	47.8	18.1	521 13 B0487067
33	47.8	18.1	522 13 B0079594
34	47.8	18.1	534 14 C0885925
35	47.8	18.1	535 13 B0093574
36	47.8	18.1	539 13 B0093535
37	47.8	18.1	539 13 B0093761
38	47.8	18.1	543 13 B0037824
39	47.8	18.1	546 12 B0048476
40	47.8	18.1	546 13 B0093534
41	47.8	18.1	549 13 B0817923
42	47.8	18.1	549 13 B0092494
43	47.8	18.1	551 14 C0605096
44	47.8	18.1	553 13 B0635815
45	47.8	18.1	554 13 B0779073

## REFERENCES

RESULT 1  
BU449043  
LOCUS 756 bp mRNA linear EST 29-NOV-2002  
DEFINITION 603765817F1 CSEQRN13 Gallus gallus CDNA clone C0828249 5' mRNA  
ACCESSION BU449043  
VERSION BU449043.1 GI:25938354  
KEYWORDS EST  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianidae; Gallus;  
1: bases to 756  
Bovatta, P. E., Sanz-Buqueiro, J., Overton, I. M., Butt, D. W., Bosch, E.,  
Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.,  
A Comprehensive Collection of Chicken cDNAs  
JOURNAL COMPANION 12 (22), 1965-1969 (2002)  
MEDLINE 2235534  
PUBMED 12445392  
COMMENT (Contract) Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology (UMIST)  
PO Box 88, Manchester, M60 1GD, UK  
Tel: 01612008930  
Fax: 01612360409

FEATURES  
Source

Email: Simon.Hubbard@wisc.ac.uk.

Location/Qualifiers  
1..756

/organism="Gallus gallus"  
/mol\_type="RNA"  
/strain="layer"  
/db\_xref="taxon:9031"  
/clone="CHEST69:1b6"

/sex="Female"  
/dev\_stage="adult"

/lab\_host="DH:03"

/note="Organ: ovary; Vector: pBluescript II KS(+); Site:1;  
ECORI; Site:2: NotI; Vector: pBluescript II KS(+); Site:1;  
ECORI; Site:2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunt-ended, ligated to NotI adapters, digested with EcoRI  
, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 793, except that a significantly longer  
reannealing hybridization was used."

BASE COUNT  
ORIGIN

142 a 234 c 238 g 142 t

Query Match  
Best Local Similarity 53.5%; Pred. No. 0.4; Length 756;  
Matches 107; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

CY 9 GGTGGAGATCAAGCCCAAGCCAGCCAGCTCACCACCCGAGCGCGTGAAGCCCAAGG 58  
DB 169 GGTGGAGATCAAGCCCAAGCCAGCCAGCTCACCACCCGAGCGCGTGAAGCCCAAGG 226  
CY 69 CAAGCCCTACAGCCCTGACACGACGCTGCTACGCTGACCTGACCTGACCTGACCTG 128  
DB 229 CCAAGCTCATCTCTCTGTGTGAAGCAAGGTTTGTGTGAAGCAAGGTTTGTGTGAAG 282  
CY 129 GGTGGATGAGATCAAGCCCAAGCCAGCCAGCTCACCACCCGAGCGCGTGAAGCCCAAGG 188  
DB 289 GCAAGCGGGCGCTGAGAGTGAAGGATTTTCGGGAGCGCTTACCCCTGCTTCCGCCCTCA 348  
CY 159 GCAAGCTGAGAGCTTCAAG 208  
DB 349 GTCATCCAGCTTCCAGG 368

RESULT 2  
LOCUS  
DEFINITION  
AM269906  
X57302.X1 Soares NFL T\_GRC S1 Homo sapiens cDNA clone  
IMAGE:2815297.3 Similar to TR:073884 073884 PUTATIVE PHOSPHATASE.  
/contains TARI.13 TARI repetitive element; mRNA sequence.  
AM269906  
AM269906 GI:6556936  
EST.  
Homo sapiens (human)  
SOURCE  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 489)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
JOURNAL  
COMMENT

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rcmail.nih.gov  
This clone is available royalty-free through LNC; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand

FEATURES  
Source

Seq primer: 43UP from Gdbco  
High quality sequence stop: 445.  
Location/Qualifiers  
1..489

/organism="Homo sapiens"  
/mol\_type="RNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2815297"  
/lab\_host="DH:08"

/lab\_host="DH:08"

/note="Organ: pooled; Vector: pTZ19-Pac (Pharmacia) with  
a modified polylinker; Site:1: Not I; Site:2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBL3W, testis NHT, and B-cell  
NCI-CCAP GCE) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
1 M.A.G.E. clones 297483-302087, 682632-697239,  
726408-728711, and 728396-731399. Subtraction by Benito  
Soares and M. Patricia Bonaldo."

BASE COUNT  
ORIGIN

77 a 181 c 153 g 77 t 1 others

Query Match  
Best Local Similarity 51.3%; Pred. No. 0.55;  
Matches 117; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

CY 34 CAGTTCAACACCCCGGAGCGCGTGAAGCCCGGAGCGCGTGAAGCCCAAGG 93  
DB 188 CCGTTCCACACACACGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 247  
CY 94 CTGTGATCAAGTCAAGCTGAGACGCTGAGACGCTGAGACGCTGAGACGCTGAGACG 153  
DB 248 AGGACATCACTGGCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307  
CY 154 GCGCAGCGCGCGCTGAGACGCTGAGACGCTGAGACGCTGAGACGCTGAGACGCTG 213  
DB 308 GCGCAGCGCGCGCGCTGAGACGCTGAGACGCTGAGACGCTGAGACGCTGAGACG 367  
CY 214 CAGTTGCGCGCGCTGAGACGCTGAGACGCTGAGACGCTGAGACGCTGAGACG 261  
DB 368 CCGCGCGCGCGCTGAGACGCTGAGACGCTGAGACGCTGAGACGCTGAGACG 415

RESULT 3  
LOCUS  
DEFINITION  
A1971889  
X57302.X1 NCI CGAP C18 Homo sapiens cDNA clone IMAGE:2815297.3  
Similar to TR:073884 073884 PUTATIVE PHOSPHATASE. mRNA sequence.  
A1971889  
A1971889 GI:5766715  
EST.  
Homo sapiens (human)  
SOURCE  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
2 (bases 1 to 493)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rcmail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Benito  
Soares, Ph.D. cDNA Library Arrayed by: Carista Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution: Information can be  
found through the I.M.A.G.E. Consortium/LNC at:  
www-bio.llnl.gov/bdip/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 383.

```

/obj_xref="taxon:9606"
/clone="IMAGE:674264"
/sex="male"
/gen_stage="20 week post conception fetus"
/lab_host="DH10B lambdafix in resistant"
/clone_id="Scars fetal liver spleen, INFUS.S1"
/issue=Organ: Liver and Spleen; Vector: pTUD (Pharmacia)
With a modified polylinker. Site 1: Pac I; Site 2: Eco RI.
This is a subtracted version of The original Scars fetal
liver spleen INFUS library. A cDNA was primed
with a Pac I - c133(5') primer 15'.
AATCGAGACATTATTTAAGACCTCTTTTCTTTTCTTTT 3').
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTUD vector. Library
went through one round of normalisation. Library
constructed by Bento Soares and X-Farina Benaldi."

```

REFERENCE NCIBases 1 to 4277  
 URL: http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS NCI-CGAP National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsabds@mai.nih.gov  
 Issue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D. CGA Library Preparation: M. Genro  
 Soares, Ph.D. CGA Library Arrayed by: Christa Frange, the  
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
 Genome Sequencing Center  
 Gene distribution: NCI-CGAP gene distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq. primer: -40bp from 3' end  
 High quality sequence: seqc 121.  
 Location/Qualifiers  
 1. 427  
 /organism="Homo sapiens"

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Ekayoyota; Mezazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 :bases : to 518)
TITLE	NCI-CCAF <a href="http://www.ncbi.nlm.nih.gov/ncicaf">http://www.ncbi.nlm.nih.gov/ncicaf</a> .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CCAF),
COMMENT	Tumor Gene Index
Unpublished	Contact: Robert Strausberg, Ph.D.
	Email: ccgaps-r@mail.nih.gov
	This clone is available royalty-free through LNC / contact the
	IMAGE Consortium ( <a href="http://infoimage.llnl.gov">infoimage.llnl.gov</a> ) for further information.
	Insert Length: 1026 Std Error: 0.00
	Seq primer: -40ml3 fwd: ET from Amerham
FEATURES	High quality sequence strop: 436.
Location/Qualifiers	
Source	1..518
	/organism="Homo sapiens"
	/mol_type="mRNA"





Query Match	18.64	Score 49.2	DB 13	Length 847
Best Local Similarity	53.38	Pred. No. 1.1		
Matches 105	Conservative	0	Mismatches 93	Indels 0
			Gaps	0

QY 1 TGAAGATCAAGCCGACGACGAGGCGCAGTGCATACACACCCGAGCGCGGTAGCGCGCAAGAGCA 70  
Db 1 TGGAGGACAAACCCCAACACCAACACAGTGGCCGAGCTACGGGAGTTTATCTCTGGCTTCC 60  
QY 71 AGCCCTTACAGCCTGAAAGAGCAGCTGTGTAAGTGAACCTGGGCTAACGATACACCCGTTGC 135  
Db 61 AGCTCATCGTCTCGCTGAAAGCAACAGTTTTTTGTGAAAGAGGCCAACCTGAACCTGGTGC 120  
QY 131 TGGTGAAAGATCAACCTCTGACAGAGGCGACACCCGCTACGCGCCCGGAGCTGAACCGCTGC 190  
Db 121 AGCGGCGGCTGAAGCTGAAGGATTTTGGGAGACCTACCCCGCTTCCCCCATCAAGT 180  
QY 191 ACCGTAGACAGCTTCAGG 208  
Db 181 CCATCCACGTTCTCCAGG 198

RESULT_12	
B2566686/c	
LOCUS	B2566686
DEFINITION	945 bp DNA linear GS: 17-DEC-2003
ACCESSION	B2566686
VERSION	B2566686.1
KEYWORDS	GS: 17-DEC-2003
SOURCE	Pseudomonas aeruginosa

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 945); Spencer, D. H., Raymond, C. K., Smith, S. E., Sims, E. E., Hastings, V., Barns, J. L., Kaul, R. and Olsen, M. V.		Whole-Genome Sequence Variation among Multiple Isolates of <i>Pseudomonas aeruginosa</i>	12541	
		<i>G. Bacteriol.</i> , 120(2), in press		
		Contact: Chris K. Raymond		

FEATURES

Location/Qualifiers

```

organism="Pseudomonas aeruginosa"
mol_type="genomic DNA"
strain="2-164"
db_xref="taxon:287"
clone="pacS2-164.6521"
clone_1b="pacS2-164"
note="1:clinical isolate 2-164 M30:e Genomic Shotgun library"

```

BASE COUNT	156	a	307	c	323	g	158	e	others
ORIGIN									

```
Query Match      18.6%; Score 49.2; DB 29; Length 945;
Best Local Similarity 51.4%; Pred. No. 1.1;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0.
```

QY 3 GTTACACACCCGGAGGGCGGTAGACCGGCGACGAGGGACGCGCAGACCTGAAACGACACGCT 95  
DB 446 GCTGGGACCCCGGTTCAACGTGGCGCGCTGGACACGCACTGTCGGGTGACCTTGAAGCG 397  
QY 96 GTCTTACCTGGACCTGGGCGAACGATACCCCTGTCGTGTAAGATCACTCTGGACGAGAG 155  
DB 386 TGGCGGGTGAAGTGTGGACCTCTGGGCTTACGGGGCAAGCGCCAGTGTGGCTCTGGCGGG 327  
QY 156 CGACGCGCGCTTACGCGCCCGGCGCTGTACACCGGTGACCTGAGCAGGCTTCAAGGTGCGGCCA 215

Db 326 CAGCAGGTCGATGCGCATTTGCGGCAACTATGACCGCTGGCCAAAGTCGATGCGGCAC 267  
 Qy 215 GTTCCGCACTGATGATCGACCGCTGCGCTGGTGCACCG 257  
 Eb 256 GCGCATGCGCTGCAAAAGACGACCGCGCTGCTTCGAGCGCAC 225

RESULT 13	974 bp	MEMA	1: near	EST: 29-MAY-2002
LOCUS	603255264F1	CSECB505	Gallus gallus	CDNA clone CSECB505.4 5', MEMA
DEFINITION	sequence.			
ACCESSION	U0412619			
VERSION	U0412619.1	GI:25305293		
KEYWORDS	EST			
SOURCE	Gallus gallus (chicken)			
ORGANISM	Gallus gallus			

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	INDEXED	COMMENT
2 (bases 1 to 974)	Boardman, P. J., Sanz-Ezquerro, J., Overton, I. M., Gurr, E. N., Bosch, E. Fong, W. T., Tinkle, C., Brown, W. R., Wilson, S. A. and Hubbard, S. J.	A Comprehensive Collection of Chicken CDNAS	Curr. Biol. 12 (22), 1965-1969 (2002)	22335534	12441392	Contact: Simon Hubbard

### FEATURES

- Location/Qualifiers

```

1097      .organism = "Homo sapiens"  

1098      .mol_type = "mRNA"  

1099      .strand = "+"  

1100      .db_xref = taxon("taxid:6496")  

1101      .cdate = "2007-06-24"  

1102      .sex = "female"  

1103      .dev_stage = "adult"  

1104      .adb_host = "CHIRB"  

1105      .clone_id = "GSCORP15"  

1106      .note = "Ovary; Vector: pBluescript II KS(+); Site:  

1107      EcorI; Site: 2; Note: Modification of pBluescript II KS(+)  

1108      (Stratagene) vector to accommodate cDNA produced with the  

1109      truncated protocol (construction of uni-directionally  

1110      cloned cDNA libraries from messenger RNA for improved 3'  

1111      end DNA sequencing by Glenn RN, et al. U.S. Patent # 6,388,  

1112      624). Cut pBluescript II KS(+) with NotI and EcoR.  

1113      Ligate in double stranded adaptor containing 5'gag and  

1114      BamI sites (5'ggccgccttcaccccccgaattccgaataaaagt)  

1115      5'aactcttttcttgcatcccggtgtgcaggc)"
```

BASE COUNT	183	a	292	c	252	5	21
ORIGIN							

Query Match	18.3%	Score 49.4	DS 13	Length 974
Best Local Similarity	55.3%	Pred. NO. 1.6		
Matches 94; Conservative	0;	Mismatches 76;	Indels 0;	Gaps 0

3y GGTGGAGATCAAGCCAGAGCCAGGCCCATTTACACACCCCGSAGGCGCGCTGAGGGCTGCCA356 68  
 Db GGTGGAGAGCACCCCMACACACACAGTGTGCGGGCTACGGGCTTTATATCTGTGCGCC 593  
 Qy CAAACCCCTACAGCTGTAAAGACAGCACTGTCTACTGTGACACTGTGCAACAGATAGCCCTCT 128  
 Db CACGCTCATATGCTCTCTCTGTAAAGCAACAAAGTTTGTGCGAAAGGAGCCAACTCTCAACTGTGT 659  
 Qy GCTGTGAAGATACCTCTGAGAGAGGGCCAGCCCGCTACGCCCCCGCCCT 179



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OX nucleic - nucleic search, using sw mode!

Run on: November 7, 2003, 12:27:20 / Search time: 187 seconds

(without alignments)  
3810.971 Million cell updates/sec

Title: US-09-622-500b-3

Perfect score: 264

Sequence: : atgatacaagtgagatcaaa.....gcccgtgcccccaagtaa z64

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001.DAT.\*  
23: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.\*  
24: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.\*  
25: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	264	100.0	264	AA211645	Coliphage M13 gene
2	147.2	55.8	6971	AA226304	Sequence of phage
3	145.6	55.2	7055	AA226303	Sequence of phage
4	145.6	55.2	7783	AA226302	Sequence of phage
5	145.6	55.2	8233	AA226301	Sequence of phage
6	144	54.5	264	AA211644	Coliphage M13 gene
7	144	54.5	7084	AA251026	Modified filament
8	144	54.5	7294	AA251027	Modified filament

9	144	54.5	7294	13	AA24174	M13X421. Synthet
10	144	54.5	7294	15	AA26626	Right-half vector
11	144	54.5	7294	15	AA26630	Right-half vector
12	144	54.5	7294	15	AA26631	M13-based vector M
13	144	54.5	7294	15	AA26632	M13-based vector M
14	144	54.5	7294	15	AA26633	Left-half vector M
15	144	54.5	7294	15	AA26634	Left-half vector M
16	144	54.5	7294	15	AA26635	Left-half vector M
17	144	54.5	7294	15	AA26636	Left-half vector M
18	144	54.5	7294	15	AA26637	Left-half vector M
19	144	54.5	7294	15	AA26638	Left-half vector M
20	144	54.5	7294	15	AA26639	Left-half vector M
21	144	54.5	7294	15	AA26640	Left-half vector M
22	144	54.5	7294	15	AA26641	Left-half vector M
23	144	54.5	7294	15	AA26642	Left-half vector M
24	144	54.5	7294	15	AA26643	Left-half vector M
25	144	54.5	7294	15	AA26644	Left-half vector M
26	144	54.5	7294	15	AA26645	Left-half vector M
27	144	54.5	7294	15	AA26646	Left-half vector M
28	144	54.5	7294	15	AA26647	Left-half vector M
29	144	54.5	7294	15	AA26648	Left-half vector M
30	144	54.5	7294	15	AA26649	Left-half vector M
31	144	54.5	7294	15	AA26650	Left-half vector M
32	144	54.5	7294	15	AA26651	Left-half vector M
33	144	54.5	7294	15	AA26652	Left-half vector M
34	144	54.5	7294	15	AA26653	Left-half vector M
35	144	54.5	7294	15	AA26654	Left-half vector M
36	144	54.5	7294	15	AA26655	Left-half vector M
37	144	54.5	7294	15	AA26656	Left-half vector M
38	144	54.5	7294	15	AA26657	Left-half vector M
39	144	54.5	7294	15	AA26658	Left-half vector M
40	144	54.5	7294	15	AA26659	Left-half vector M
41	144	54.5	7294	15	AA26660	Left-half vector M
42	144	54.5	7294	15	AA26661	Left-half vector M
43	144	54.5	7294	15	AA26662	Left-half vector M
44	144	54.5	7294	15	AA26663	Left-half vector M
45	144	54.5	7294	15	AA26664	Left-half vector M

#### ALIGNMENTS

RESULT 1	AA211645	standard; DNA: 264 bp.
AC	AA211645:	
DT	19-NOV-1999	(first entry)
DE	Coliphage M13 gene 5 synthetic nucleic acid sequence.	
EE	Gemini virus resistance; ssDNA virus; ssDNA-binding protein; IVV;	
EE	Infectious virus; Coliphage; gene 5 protein; viral infection;	
EE	Plant virus; plant protection; ss.	
EE	Coliphage sp.	
EE	Synthetic.	
EE	Key	Location/Qualifiers
EE	CD8	1..264
EE		/+tag= a
EE	MO3945101-A1.	
EE	10-SEP-1999.	
EE	03-MAR-1999:	99WO-US04716.
EE	03-MAR-1999:	98US-0076627.
EE	(SCRIPT )	SCRIPTS RES INST.

PI Padidar M, Beachy RN, Fauquet CM;  
XX WPI; 1999-551037/46.  
DR P-PSDB; AAY3369.  
XX  
PT Producing plants resistant to single stranded DNA virus, using a single  
PT stranded DNA-binding protein of the Inoviridae virus family  
XX  
PS Claim 9; Page 61; 69pp; English.  
XX  
CC The invention relates to a method for producing in a plant geminivirus  
CC resistance to a single stranded (ss) DNA virus. The method comprises  
CC introducing a gene capable of expressing a ssDNA-binding protein of the  
CC Inoviridae virus (IV) family into the plant. The IV family virus is  
CC selected from novoviruses or Plectrovirus genera where the novovirus genus  
CC phage, Pseudomonas phage, Vibrio phage or Xanthomonas phage species. The  
CC ssDNA-binding protein is a Coliphage coat protein or a gene 5 protein  
CC and especially a Coliphage M13 gene 5 protein. The ssDNA-binding protein  
CC of the IV family interferes with virus spread during the infection  
CC process of plant viruses of the ssDNA type. By inhibiting virus spread,  
CC the virus infection is reduced and/or blocked, thereby increasing plant  
CC resistance to the virus infection. The methods can be used for protecting  
CC plants against e.g. Bajra streak virus, Bean yellow dwarf virus, Bromus  
CC striate mosaic virus, Chickpea chlorotic dwarf virus, Chloris striate  
CC mosaic virus, Digitaria streak virus, Digitaria striate mosaic virus,  
CC maize streak virus, Miscanthus streak virus, Tobacco yellow dwarf virus,  
CC wheat dwarf virus, Beet curly top virus, Horseadish curly top virus,  
CC Tomato leafroll virus, Tomato pseudo-curly top virus, Abutilon rosette  
CC virus, Kalypna yellow mosaic virus, cassava mosaic virus, Tomato leaf  
CC curly virus, Cotton leaf curl virus, Banana bunchy top virus, Coconut  
CC foliar decay virus, Fabbean necrotic yellow virus, Milk vetch dwarf  
CC virus or Subterranean clover stunt virus. The present sequence  
CC represents a synthetic DNA sequence of Coliphage M13 gene 5 protein.  
XX  
SQ Sequence 264 BP; 54 A; 93 C; 80 G; 37 T; 0 other;  
Query Match 100.0%; Score 264; DB 20; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.7e-42;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGATCAAGGTGAGATTCAGAGCCAGCCAGCCAGTTCACCAACCGCGGCGTACG 60  
DB 1 ATGATCAAGGTGAGATTCAGAGCCAGCCAGCCAGTTCACCAACCGCGGCGTACG 60  
QY CGCCAGGCGCAAGCCCTACAGCCCTGAGAGCAAGCACTGCTGACCTGGAGCGGCAACAG 120  
DB 61 CGCCAGGCGCAAGCCCTACAGCCCTGAGAGCAAGCACTGCTGACCTGGAGCGGCAACAG 120  
QY 121 TACCCGCTGCTGTGAAGATCACTCTGAGAGAGGAGGAGCCGCTACGCCCGCGGCTG 180  
DB 121 TACCCGCTGCTGTGAAGATCACTCTGAGAGAGGAGGAGCCGCTACGCCCGCGGCTG 180  
QY 181 TACACCGTGACCTGAGAGCACTTCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
DB 181 TACACCGTGACCTGAGAGCACTTCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
QY 241 CTGCGCTGTGCTGCGCGGCAAGTAA 264  
DB 241 CTGCGCTGTGCTGCGCGGCAAGTAA 264  
RESULT 2  
AAX26304  
ID AAX26304 standard; DNA; 6971 BP.  
XX  
AC AAX26304;  
XX  
CT 21-MAY-1999 (first entry);  
XX  
DE Sequence of phage vector fpep3\_1B-IR3seq.  
XX  
KW Phage vector; fpep3\_1B-IR3seq; multimeric protein complex;

KW polypage particle; antibody Fab fragment; enzyme complex; de  
XX  
CS Synthetic;  
XX  
PN WC3906587-A2.  
XX  
XX 11-FEB-1999.  
XX  
XX  
XX 03-AUG-1998; 98WO-EP04836.  
XX  
XX 01-AUG-1997; 97EP-0113319.  
XX  
PA (MORP-1) MORPHOSYS GDS PROTEINOPTIMIERUNG NRB.  
XX  
XX Ge L, Haag V, Rudert F;  
XX  
DR WPI; 1999-151811/13.  
XX  
PT Identifying combinations of nucleic acids that encode multimeric  
PT protein complexes - based on new polypage particles that display  
PT both components simultaneously, particularly for selecting antibody  
PT Fab fragments or enzyme complexes, does not require a recombination  
PT event  
XX  
XX Claim 14; Fig 4; 72pp; English.  
XX  
CC The present sequence represents the phage vector fpep3\_1B-IR3seq, which  
CC was used in the course of the invention. The specification describes a  
CC method for identifying combinations of nucleic acids that encode  
CC multimeric protein complexes. The method is based on polypage  
CC particles that display both components simultaneously, particularly  
CC for selecting antibody Fab fragments or enzyme complexes. In addition,  
CC the method does not require a recombination event, since polypages  
CC co-package both libraries. Both members of the multimeric polypeptide  
CC complex, particularly the light and heavy chains of an Fab fragment,  
CC can be identified simultaneously.  
XX  
SQ Sequence 6971 BP; 1737 A; 1402 C; 1463 G; 2369 T; 0 other;  
Query Match 55.8%; Score 147.2; DB 20; Length 6971;  
Best Local Similarity 72.3%; Pred. No. 5.1e-20;  
Matches 192; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 1 ATGATCAAGGTGAGATTCAGAGCCAGCCAGCCAGTTCACCAACCGCGGCGTACG 60  
DB 4489 ATGATCAAGGTGAGATTCAGAGCCAGCCAGCCAGTTCACCAACCGCGGCGTACG 60  
QY 61 CGCCAGGCGCAAGCCCTACAGCCCTGAGAGCAAGCACTGCTGACCTGGAGCGGCAACAG 120  
DB 4549 CGCCAGGCGCAAGCCCTACAGCCCTGAGAGCAAGCACTGCTGACCTGGAGCGGCAACAG 120  
QY 121 TACCCGCTGCTGTGAAGATCACTCTGAGAGAGGAGGAGCCGCTACGCCCGCGGCTG 180  
DB 121 TACCCGCTGCTGTGAAGATCACTCTGAGAGAGGAGGAGCCGCTACGCCCGCGGCTG 180  
QY 181 TACACCGTGACCTGAGAGCACTTCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
DB 181 TACACCGTGACCTGAGAGCACTTCAGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 240  
QY 241 CTGCGCTGTGCTGCGCGGCAAGTAA 264  
DB 4729 CTGCGCTGTGCTGCGCGGCAAGTAA 264  
RESULT 3  
AAX26303  
ID AAX26303 standard; DNA; 7055 BP.  
XX  
AC AAX26303;  
XX  
CT 21-MAY-1999 (first entry);  
XX  
DE Sequence of phage vector fjun\_1B.

```

XX XX Phage vector; fjun; IB; multimeric protein complex; polyphage particle;
KM antibody Fab fragment; enzyme complex; ds.
XX Synthetic.
XX MO9906587-A2.
XX
XX 11-FEB-1999.
XX
XX 03-AUG-1998; 98WO-EP04836.
XX
XX 01-AUG-1997; 97EP-0113319.
XX
XX (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.
XX
XX Ge L, Ilag V, Rudert F;
XX
XX MPI; 1999-153811/13.
XX
XX PT Identifying combinations of nucleic acids that encode multimeric
XX PT protein complexes - based on new polyphage particles that display
XX PT both components simultaneously, particularly for selecting antibody
XX PT Fab fragments or enzyme complexes, does not require a recombination
XX PT event.
XX
XX Example 2; Fig 3; 71pp; English.
XX
XX CC The present sequence represents the phage vector fjun; IB, which was
XX CC used in the course of the invention. The specification describes a
XX CC method for identifying combinations of nucleic acids that encode
XX CC multimeric protein complexes. The method is based on polyphage
XX CC particles that display both components simultaneously, particularly
XX CC for selecting antibody Fab fragments or enzyme complexes. In addition,
XX CC the method does not require a recombination event, since polyphages
XX CC co-package both libraries, both members of the multimeric polypeptide
XX CC complex, particularly the light and heavy chains of an Fab fragment,
XX CC can be identified simultaneously.
XX
XX SO Sequence 7055 BP; 1769 A; 1433 C; 1492 G; 2357 T; 4 other;
XX
XX Query Match: 55.2%; Score 145.6; DB 20; Length 7055;
XX Best Local Similarity 72.0%; Pred. No. 1,2e-19;
XX Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
XX
XX QY 1 ATGATCAAGGTGAGATCAAGCCGACGAGCCGACGACGACGAGCGGCGTGGAGC 60
XX DB 843 ATGATTAAGTTGAATTAACCGCTCAAGGCGAAATTACACCGTTGAGTCTTCT 902
XX QY 61 CCGCAGGCGAACCCCTACAGCTGAAACGAGCGAGCTGTGTAAGTGGACCTGGCAAGAG 120
XX DB 903 CGTCAGGCGAACCTTATTCAGTGAATGAGCGAGCTTGTACGTTGATTTGGGTAATGAA 962
XX QY 121 TACCCGCTGCTGGTGAAGATCAACCCCTGAGAGGAGCGGACCGCTACGCGCGGCTTG 190
XX DB 963 TATCCGCTGCTGTCAAGATTAAGTCTGCAAGAGAGTGAAGCGAGCTATGCGCTGCTG 1022
XX QY 181 TACAGCGTGAAGCTGAGCAAGCTTCAAGTGGCGCAAGTTGCGAGCGTGAATGACAGCG 240
XX DB 1023 TACAGCGTGAAGCTGAGCAAGCTTCAAGTGGCGCAAGTTGCGAGCGTGAATGACAGCG 1082
XX QY 241 CTGCGCTGCTGGTGGCGCGCAAGTAA 264
XX DB 1083 CTGCGCTGCTGGTGGCGCGTAAAGTAA 1106
XX
XX RESULT 4
XX AAX26302
XX ID AAX26302 standard; DNA; 7783 BP.
XX AC AAX26302;
XX XX
XX 21-MAY-1999 (first entry)

```

```

XX XX Sequence of phage vector fjun; IB.
XX DE Phage vector; fjun; IB; multimeric protein complex; polyphage particle;
XX XX antibody Fab fragment; enzyme complex; ds.
XX KM Synthetic.
XX OS MO9906587-A2.
XX
XX 11-FEB-1999.
XX
XX 03-AUG-1998; 98WO-EP04836.
XX
XX 01-AUG-1997; 97EP-0113319.
XX
XX (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.
XX
XX Ge L, Ilag V, Rudert F;
XX
XX MPI; 1999-153811/13.
XX
XX PT Identifying combinations of nucleic acids that encode multimeric
XX PT protein complexes - based on new polyphage particles that display
XX PT both components simultaneously, particularly for selecting antibody
XX PT Fab fragments or enzyme complexes, does not require a recombination
XX PT event.
XX
XX Example 2; Fig 3; 71pp; English.
XX
XX CC The present sequence represents the phage vector fjun; IB, which was
XX CC used in the course of the invention. The specification describes a
XX CC method for identifying combinations of nucleic acids that encode
XX CC multimeric protein complexes. The method is based on polyphage
XX CC particles that display both components simultaneously, particularly
XX CC for selecting antibody Fab fragments or enzyme complexes. In addition,
XX CC the method does not require a recombination event, since polyphages
XX CC co-package both libraries, both members of the multimeric polypeptide
XX CC complex, particularly the light and heavy chains of an Fab fragment,
XX CC can be identified simultaneously.
XX
XX SO Sequence 7783 BP; 1920 A; 1649 C; 1667 G; 2543 T; 4 other;
XX
XX Query Match: 55.2%; Score 145.6; DB 20; Length 7783;
XX Best Local Similarity 72.0%; Pred. No. 1,2e-19;
XX Matches 190; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
XX
XX QY 1 ATGATCAAGGTGAGATCAAGCCGACGAGCCGACGACGAGCGGCGTGGAGC 60
XX DB 843 ATGATTAAGTTGAATTAACCGCTCAAGGCGAAATTACACCGTTGAGTCTTCT 902
XX QY 61 CCGCAGGCGAACCCCTACAGCTGAAACGAGCGAGCTGTGTAAGTGGACCTGGCAAGAG 120
XX DB 903 CGTCAGGCGAACCTTATTCAGTGAATGAGCGAGCTTGTACGTTGATTTGGGTAATGAA 962
XX QY 121 TACCCGCTGCTGGTGAAGATCAACCCCTGAGAGGAGCGGACCGCTACGCGCGGCTTG 180
XX DB 963 TATCCGCTGCTGTCAAGATTAAGTCTGCAAGAGAGTGAAGCGAGCTATGCGCTGCTG 1022
XX QY 181 TACAGCGTGAAGCTGAGCAAGCTTCAAGTGGCGCAAGTTGCGAGCGTGAATGACAGCG 240
XX DB 1023 TACAGCGTGAAGCTGAGCAAGCTTCAAGTGGCGCAAGTTGCGAGCGTGAATGACAGCG 1082
XX QY 241 CTGCGCTGCTGGTGGCGCGCAAGTAA 264
XX DB 1083 CTGCGCTGCTGGTGGCGCGTAAAGTAA 1106
XX
XX RESULT 5
XX AAD48845
XX ID AAD48845 standard; DNA; 8233 BP.
XX AC AAD48845;
XX XX

```



Db	61	CGTGGGCGAAGCCTTATTCAGTGAATGACGAGCTTGTACGTGATTTGGGTAA	120
Gy	121	TACCCGCTGTGTGTAAGATCACCCGAGCAGGGCCAGCCGCGCTACGCCCGCGCTG	130
Db	122	TATCGGCTCTTGTCAAGATTAATCTGTATGAGATCAGCCAGCCCTATGCCCTCGCTCG	130
Gy	181	TACACCGCTGCACCTGAGCAGCCTTCAGAGTCGCGCAGTTGGGAGCGCTGATGATGATCGC	240
Db	181	TACACCGTTGATCTGTCTCTTTCAGAGTTGGTCACTTGGTCCCTTAATGATGACCGT	240
Gy	241	CTGAGCTGTGGCCGCGCAATGA	264
Db	241	CTGGGCTCTTCCGCTAAGTAA	264

QY 61 CCGCAGACGGCAAGCCCTTACAGCCCTAACAGGAGACCTGGCTACAGTACACCTCTGGGCAAGAG 122  
DB 1772 CGTCAAGGCAAGCCCTTTTACCTAAATGAGGAGCTTTCTTACGTTGATTTGGGTATATAA 1833  
QY 121 TACCCCGGTGCTGGTGAAGATACAGCCTGGACAGAGAGACAGCTCGGCTAACGATGCGCGAGCTTG 180  
DB 1832 TATCCCGGCTCTGTGTCAAGATTACTCTTGATTAAAGTTCAAGCAGATCTAAGCGCGCTGGCTG 1533  
QY 181 TACACCGTGCACCTGAGAGACCTTCAAGGTGGGCGACGTTGGGACGCGCTATATATCGAACCGC 240  
DB 1832 TACACCGTGCATCTGTCTCTTTCAAAAGTTGTCAGTTGGCTTATGATATGATTAACCGT 1953  
QY 241 CTGGCGCCTGGTGGCCCGGCAATAA 264  
DB 1952 CTGGCGCCTCGTTTGGGCTAAAGTAA 1975

Query Match 54.5%; Score 144; DB 13; Length 7294;  
 Best Local Similarity 71.6%; Pred. No. 2.5e-19;  
 Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGGAGATCAAGCCAGCCAGGCGGAGCTTACACCCGACGCGCGCTAGAC 60  
 DB 643 ATGATTAAGTTAAATTAACCATCTCAAGCCCAATTACACTGCTGTCTGTCTCT 902  
 QY 61 GCGCAGGCGCAAGCCCTACAGCCTGAACGACAGCTGTCTAGCTGAGACTGCGCAACAG 120  
 DB 903 CGTCAGGCGCAAGCCCTATTCAGTAAGAGCTGTGTAGCTGATTTGGGTAAATGAA 962  
 QY 121 TACCCGCTGTGTGTAAGATCAACCTGACAGAGGCGACGCGCCCTACGCGCGCGCTG 180  
 DB 963 TATCCGCTGTGTGTAAGATTAAGAGCTGTGTAGAGGCTACAGCCCAATGCGCTGTCTG 1022  
 QY 181 TACACGCTGGACCTGAGACAGCTTCAGAGCTGCGACATTCGCGAGCCTGATGATGACAGC 240  
 DB 1023 TACACGCTGATCTGTCTCTCTTCAAGAGTGGTCAAGTGGTGGTCCCTTATGATGACCGT 1082  
 QY 241 CTGCGCCTGTGTGCGCGCGCAAGTAA 264  
 DB 1083 CTGCGCCTGTGTGCGCGCTAAGTAA 1106

RESULT 9  
 AAQ24174  
 ID AAQ24174 standard; DNA: 7294 BP.  
 AC AAQ24174;  
 DT 19-JAN-1993 (first entry)  
 XX M31X421.  
 DE M31X421.  
 XX M13; cloning vectors; gvt11; pseudo-wild type; coat protein; RBS.  
 KM LacZ; ss.  
 XX Synthetic.  
 OS  
 XX W09206176-A.  
 XX 16-APR-1992.  
 XX 27-SEP-1991; 91MO-US07.41.  
 XX 28-SEP-1990; 90US-0590664.  
 PA (IXSY-) IXSYS INC.  
 PI Huse WD;  
 PS WPI; 1992-150863/18.  
 CC The sequences given in AAQ24173-5 are vectors which allow the expression  
 CC of random peptides on the surface of M13. The vectors are produced by  
 CC combining separate vectors into a single larger vector. This system  
 CC produces random oligonucleotides functionally linked to expression  
 CC elements and to gvt11. These vectors encode a pseudo-wild type gvt11  
 CC product. This gene encodes the wild-type M13 gvt11 amino acid sequence  
 CC but has been changed at the nucleotide level to reduce homology  
 CC recombination with the wild-type gvt11 contained in these vectors.  
 CC The wild-type gvt11 is present to ensure that at least some functional,  
 CC non-fusion coat protein is produced. They also contain the expression  
 CC elements for the peptide fusion proteins. A ribosome binding site and

CC LacZ promoter/operator elements are present for transcription and  
 CC translation of the peptide fusion proteins. Various restriction  
 CC sites are present for the cloning of random peptides.  
 XX Sequence 7294 BP; 1796 A; 1527 C; 1553 G; 2418 T; 0 other;

QY 1 ATGATCAAGGTGGAGATCAAGCCAGCCAGGCGGAGCTTACACCCGACGCGCGCTAGAC 60  
 DB 643 ATGATTAAGTTAAATTAACCATCTCAAGCCCAATTACACTGCTGTCTGTCTCT 902  
 QY 61 GCGCAGGCGCAAGCCCTACAGCCTGAACGACAGCTGTCTAGCTGAGACTGCGCAACAG 120  
 DB 903 CGTCAGGCGCAAGCCCTATTCAGTAAGAGCTGTGTAGCTGATTTGGGTAAATGAA 962  
 QY 121 TACCCGCTGTGTGTAAGATCAACCTGACAGAGGCGACGCGCGCTACGCGCGCGCTG 180  
 DB 963 TATCCGCTGTGTGTAAGATTAAGAGCTGTGTAGAGGCTACAGCCCAATGCGCTGTCTG 1022  
 QY 181 TACACGCTGGACCTGAGACAGCTTCAGAGCTGCGACATTCGCGAGCCTGATGATGACAGC 240  
 DB 1023 TACACGCTGATCTGTCTCTCTTCAAGAGTGGTCAAGTGGTGGTCCCTTATGATGACCGT 1082  
 QY 241 CTGCGCCTGTGTGCGCGCGCAAGTAA 264  
 DB 1083 CTGCGCCTGTGTGCGCGCTAAGTAA 1106

RESULT 10  
 AAQ66626  
 ID AAQ66626 standard; DNA: 7294 BP.  
 AC AAQ66626;  
 DT 25-MAR-2003 (updated;  
 DT 20-JAN-1995 (first entry)  
 XX Right-half vector M13IX42.  
 XX vector M13IX42; right half vector; randomised oligonucleotides;  
 KM surface expression; random peptide; constrained secondary structure;  
 KM pseudo wild-type M13 gene VIII; ds.  
 XX Synthetic.  
 OS  
 XX W09411496-A1.  
 XX 26-MAY-1994.  
 XX 09-NOV-1993; 93MO-US-0650.  
 XX 10-NOV-1992; 92US-0978893.  
 PA (IXSY-) IXSYS INC.  
 PI Huse WD;  
 PS WPI; 1994-181498/22.  
 CC Cells expressing oligonucleotide(s) having random codon  
 CC sequences - are used for producing soluble peptide(s) having a  
 CC constrained secondary structure in solution.  
 XX Example 1; Page 76-82; 152pp; English.  
 XX M13IX42 was constructed to harbour the right-half populations of  
 CC randomised oligonucleotides. M13mp8 was the starting vector which  
 CC was modified to contain, in addition to the encoded wild-type M13  
 CC gene VIII all ready present in the vector: a pseudo-wild-type M13  
 CC gene VII; sequence with a stop codon placed between it and an EcoRI-



```

CC amber codon between the EcoR-SacI cloning site and the pseudo-wild
CC type gene with sequence was removed. This change ensures that all
CC expression off the late promoter produces a peptide-gene with
CC fusion protein. The vector was used for the construction of 1987
CC half oligonucleotide libraries.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 7294 BF; 1787 A; 1528 C; 1553 G; 2466 T; 0 other.
SO
Query Match          54.5%; Score 144; DS 15; Length 7294;
Best Local Similarity 72.6%; Pred. No. 2,567.9;
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0.
OY      1 ATGATCAGGTGGGATCAACGCCACAGCGCCAGCTTACCACCCGCAGCCGGCTCACG 60
Dc      844 AGATTTAAAGTTGAATTMAACCATCCCAAGGCCAATTACTACTCTTCCTGGTCTTCT 902
OY      61 CGCCAGGCGAAGCCCTTACAAGCCTGTAAGAGAACACCTGTGCTACGTGACCTGGGCAACAG 120
Dc      903 CCGTAGGCGAAGCCCTTATCACTGTAAAGACACCTTTGTACGTTATTTGGGATATSA 962
OY      121 TACCCCCGCTGGTGAGATCAACCTGGAGAGAGGACACCCCGCTACGCGGCGCTG 180
Dc      963 TATCCGGGTTCTTG-CMAAGTACTCTTGATGAAGGATCAAGCCATATAAGCCCTATCTG 1022
OY      191 TACACCGCTGACCTGACAGCTTTAAGCTGGGCAAGTGGGCAGACCTGATGATGACGCG 240
Dc      1023 TACACCGCTTACATCGTCTCTTTTAAAGTTGGTCAAGTGGGTTCCTTATGATTAACGGT 1082
OY      241 CTGGCGCTGCTGCCCGCCCAAGTAA 264
Dc      1083 CTGGCGCTGCTGGCGGCTAAGTAA 1106
RESULT 12
AAV03651
ID     AAV03651 standard. DNA: 7317 BP.
XX
XX AAV03651:
XX AC      AAV03651:
XX CC      02-APR-1998 (first entry)
XX CT      M13-based vector M13IX...
XX DE      Vector construction: T-cell receptor antibody production; diagnosis;
XX KM      heteromeric receptor; immunoglobulin superfamily; therapy; circular;
XX KW      M13-based vector; ds.
XX
XX CS      Synthetic.
XX
XX PN      US5698426-A.
XX PD      16-DEC-1997.
XX XX      05-JUN-1995;    95US-0464136.
XX PR      27-SEP-1991;    91US-0767136.
XX PR      28-SEP-1990;    90US-0590215.
XX PR      13-SEP-1993;    93US-0120648.
XX PR      01-DEC-1994;    94US-0349131.
XX PR      05-JUN-1995;    95US-0464136.
XX PA      (IXSY-) IXSYS INC.
XX PI      Rise WD;
XX PT      WPI; 1996-051493/05.
XX DR      Construction of heteromeric receptor libraries - by combining
XX FT      vectors encoding different immunoglobulin fragments
XX PS      Example 1; Column 27-34; 57bp; English.
XX

```

CC This sequence represents a Y33-based vector constructed using the method  
CC of the invention, and the primers shown in AA03629-V03649. The method  
CC of the invention is for constructing a diverse population of vectors  
CC capable of expressing diverse populations of first and second  
CC polypeptides which form a diverse population of heteromeric receptors of  
CC the immunoglobulin superfamily, and comprises: (a) operationally linking  
CC to a first vector a first population of diverse DNA sequences within a  
CC first gene family encoding a diverse population of first polypeptides of  
CC the heteromeric receptor of the immunoglobulin superfamily; the first  
CC vector having two pairs of restriction sites symmetrically oriented about  
CC a cloning site; (b) operationally linking to a second vector a second  
CC population of diverse DNA sequences within a second gene family encoding  
CC a diverse population of second polypeptides of the heteromeric receptor  
CC of the immunoglobulin superfamily; the second vector having two pairs of  
CC restriction sites symmetrically oriented about a cloning site in an  
CC identical orientation to that of the first vector, where polypeptides  
CC encoded by the first and second gene families are known to form  
CC heteromeric receptors of the immunoglobulin superfamily; and (c)  
CC combining the vector sequences produced by steps (a) and (b) under  
CC conditions which allow only the operational combination of the vector  
CC sequences containing the first and second populations of diverse DNA  
CC sequences. The method is used to produce diverse populations of  
CC antibodies or T-cell receptors for diagnostic or therapeutic purposes.

SC Sequence 7317 BP; 1787 A; 1555 C; 1547 G; 2428 T; 0 other:  
Query Match 54.5%; Score 144; DB 19; Length 7317;  
Best Local Similarity 71.6%; Pred. No. 2.5e-19;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

DY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTCAACCCGAGCGAGCGTGAAC 60  
DB 843 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTCAACCCGAGCGTGAAC 60  
DY 61 CGCAGGCGAAGCCCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 903 CGCAGGCGAAGCCCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DY 121 TACCCGCTGCTGGTGAAGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 963 TATCCGCTGCTGGTGAAGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DY 181 TACACCGTGCACCTGAGACAGCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 1223 TACACCGTGCACCTGAGACAGCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DY 241 CTGCGCTGGTGGCCGCGCAAGTAA 264  
DB 1083 CTGCGCTGGTGGCCGCGCAAGTAA 1106

RESULT 13  
AA016953  
ID AA016953 standard; DNA; 7317 BP.

AC AA016953;  
XX 11-MAY-1999 (first entry)  
XX Plasmid M13X11.  
XX Heteromeric; receptor; immunoglobulin; superfamily; plasmid; primer; PCR;  
KW Bacteriophage; fusion protein; amplification; heavy chain; light chain;  
KW immune system; diagnosis; cyclic; circular; ds.  
XX Synthetic.  
XX OS  
XX US871974-A.  
XX PN  
XX 16-FEB-1999.  
XX PD  
XX 02-DEC-1994; 94US-0349131.  
XX PF  
XX

PR 27-SEP-1994; 94US-0767136  
PR 28-SEP-1990; 90US-0590219  
PR 15-SEP-1993; 93US-0120648  
PR 02-DEC-1994; 94US-0149131.  
XX  
PA (IXSY-) IXSYS INC.  
XX  
XX Huse KD;  
XX WPI: 1993-166647/14.  
XX  
XX  
XX New surface expression libraries expressing heteromeric receptors  
XX comprising cells containing vectors containing combinations of DNA  
XX sequences encoding first and second polypeptides  
XX  
XX Example 1: Fig 3A-C; 58pp; English.  
XX  
XX The invention relates to the expression of heteromeric receptor  
XX proteins, e.g. from an immunoglobulin (Ig) superfamily, in cells  
XX containing the heteromeric receptor genes on a single plasmid.  
XX Especially mentioned, the cell may be a bacteriophage, where the receptor  
XX protein are expressed as fusion proteins with the surface protein GVI.  
XX This sequence represents the plasmid M13X11 which is used for expression  
XX of the light chain proteins. The plasmid is constructed using primers  
XX AA016953-016952. The methods can be used to generate diverse populations  
XX of heteromeric receptors which mimic the natural immune system and can be  
XX used for diagnostic and therapeutic purposes.

SC Sequence 7317 BP; 1787 A; 1555 C; 1547 G; 2428 T; 0 other:  
Query Match 54.5%; Score 144; DB 20; Length 7317;  
Best Local Similarity 71.6%; Pred. No. 2.5e-19;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

DY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTCAACCCGAGCGAGCGTGAAC 60  
DB 843 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTCAACCCGAGCGAGCGTGAAC 60  
DY 61 CGCAGGCGAAGCCCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 903 CGCAGGCGAAGCCCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DY 121 TACCCGCTGCTGGTGAAGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 963 TATCCGCTGCTGGTGAAGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DY 181 TACACCGTGCACCTGAGACAGCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 1223 TACACCGTGCACCTGAGACAGCTTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DY 241 CTGCGCTGGTGGCCGCGCAAGTAA 264  
DB 1083 CTGCGCTGGTGGCCGCGCAAGTAA 1106

RESULT 14  
AA016953  
ID AA016953 standard; DNA; 7317 BP.

AC AA016953;  
XX 25-MAY-2000 (first entry)  
XX  
XX  
XX Kappa light chain library nucleotide sequence of M13X11.  
XX Bacteriophage M13 vector; prokaryotic cell; heteromeric receptor;  
KW antibody; immune system; filamentous bacteriophage; cloning;  
KW screening; coexpression; ds.  
XX  
XX Mus sp.  
XX Bacteriophage M13.  
XX OS  
XX Synthetic.  
XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 7, 2003, 12:56:12 ; Search time 194 Seconds

(without alignments)  
4342.314 Million cell updates/sec

Title: US-09-622-500b-3

Sequence: 1 atgacacaggttgagatcaaa.....gcctggcgccgcacagtaa 264

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 214354 seqs, 159547879 residues

Total number of hits satisfying chosen parameters: 4287708

Minimum DB seq length: 5

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US6C\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US6C\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	54.5	7083	10	US-09-995-396-1
2	144	54.5	7294	9	US-09-727-311-1
3	144	54.5	7294	9	US-09-727-311-5
4	144	54.5	7320	9	US-09-727-311-2
5	144	54.5	7394	9	US-09-727-311-6
6	144	54.5	7409	9	US-09-727-311-4
7	144	54.5	7445	9	US-09-727-311-3
8	144	54.5	8195	12	US-10-378-557-57
9	133.6	50.6	8171	12	US-10-378-557-58
10	54.2	20.5	2133	14	US-10-156-761-97
11	54.2	20.5	9025608	14	US-10-156-761-1
12	50.4	19.1	759	12	US-10-329-386-2795
13	46.6	17.7	972	12	US-10-259-165-321
14	46.4	17.6	1242	14	US-10-156-761-5818
15	46.4	17.6	9025608	14	US-10-156-761-1
16	45.6	17.3	2010	12	US-10-414-637-9

17	45.6	17.3	2010	13	US-10-329-386-2795
18	45.4	17.2	722	14	US-10-156-761-6008
19	45.4	17.2	1782	14	US-10-156-761-5623
20	45.2	17.1	1479	12	US-09-367-454-67
21	45	17.0	1458	14	US-10-156-761-6684
22	44.4	16.8	1479	12	US-09-367-454-63
23	44.4	16.8	1506	12	US-09-367-454-66
24	44.4	16.8	2242	12	US-10-190-435-20
25	44.4	16.8	2742	12	US-10-190-435-57
26	44.4	16.8	2742	12	US-10-190-305A-15
27	44.4	16.8	2742	12	US-10-190-305A-81
28	44.4	16.8	2739	12	US-10-241-009-18
29	44.4	16.8	2739	12	US-10-241-009-53
30	44.4	16.8	2739	12	US-10-190-435-18
31	44.4	16.8	2739	12	US-10-190-435-13
32	44.4	16.8	2739	12	US-10-190-305A-11
33	44.4	16.8	2799	12	US-10-190-305A-74
34	44.4	16.8	3162	12	US-10-190-435-18
35	44.4	16.8	3205	12	US-10-241-009-17
36	44.4	16.8	3205	12	US-10-190-435-17
37	44.4	16.8	3482	12	US-10-241-009-16
38	44.4	16.8	3482	12	US-10-241-009-15
39	44.4	16.8	3482	12	US-10-190-435-15
40	44.4	16.8	3531	12	US-10-190-435-13
41	44.4	16.8	3537	12	US-10-190-435-13
42	44.4	16.8	3537	12	US-10-190-435-15
43	44.4	16.8	3544	12	US-10-241-009-12
44	44.4	16.8	3544	12	US-10-241-009-14
45	44.4	16.8	3564	12	US-10-190-435-13

#### ALIGNMENTS

RESULT:  
US-09-995-396-1  
Sequence 1, Appl1  
Patent No. US20020446827A1  
GENERAL INFORMATION:  
APPLICANT: Kary, Arday S.  
TITLE OF INVENTION: MODULATION OF POLYPEPTIDE DISPLAY ON NUCLEOTIDE  
FILE REFERENCE: SCR22215  
CURRENT APPLICATION NUMBER: US/09/995,396  
CURRENT FILING DATE: 2001-11-27  
PRIOR APPLICATION NUMBER: 09/552,834  
PRIOR FILING DATE: 2001-04-09  
PRICE APPLICATION NUMBER: 60/096,326  
PRIOR FILING DATE: 1998-08-12  
PRIOR APPLICATION NUMBER: 09/98,839  
PRIOR FILING DATE: 1998-12-24  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 7083  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: a modified  
OTHER INFORMATION: filamentous phage  
US-09-995-396-1

Query Match: 54.5% Score 144; DB 10; Length 7083;  
Best Local Similarity 71.6%; Pred. 1.7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
CY : ATATCAAGTGTGAGATCAAGCCACCCAGGCGGCACTTACACCCGCGCGGTGAC 60  
DB : ATATTAAGTGTGAGATCAAGCCACCCAGGCGGCACTTACACCCGCGCGGTGAC 1770  
CY : CGCAGGCGAGCGCTTACAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 120  
DB : CGCAGGCGAGCGCTTACAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1771  
CGCAGGCGAGCGCTTACAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1830

QY 121 TACCCGCTGCTGGAAGATACCCGACAGAGCCAGCCGCTTACCCGCTG 186  
DB 1831 TATCCGCTGCTGGAAGATACCCGACAGAGCCAGCCGCTTACCCGCTG 1892  
QY 181 TACACCGTGACCTGAGAGCTTCAAGTGGGAGCTGCGACGCTGATGATGACGCG 240  
DB 1891 TACACCGTGATCTGCTCTTTGAAAGTGGTGAGTGGCTTCCCTATGATGACGCT 1938  
QY 241 CTGCGCTGTGCTCCGCGCAAGTAA 264  
DB 1951 CTGCGCTGTGCTCCGCGCAAGTAA 1974

## RESULT 2

US-09-727-311-1  
Sequence 1, Application US/39727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William E.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
NUMBER OF SEQUENCES: 61  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 200C  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 9072  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7294 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-1

Query Match 54.5%; Score 144; DB 9; Length 7294;  
Best Local Similarity 71.6%; Pred. No. 1,76-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGCTGAGATCAAGCCAGCCAGCTGCTACGCTGAGACCTGGCAAGC 50  
DB 843 ATGATTAAG-TGAATTAACCATCTCAAGCCCAATTACTCTTCTGTGGTTCT 902  
QY 61 CGCCAGGCGAAGCCCTACAGCCTTAAGCAGAGCTGCTACGCTGAGACCTGGCAAGC 120  
DB 963 CGTCAGGCGAAGCCCTATTACCTGAATGAGCAGCTTTGTACGTTGATTTGGGTAAGAA 962  
QY 121 TACCCGCTGCTGGAAGATCAAGCTGAGAGGCGCAGCCGCTTACGCTGAGCTG 185  
DB 963 TATCCGCTGCTGGAAGATCAAGCTGAGAGGCGCAGCCGCTTACGCTGAGCTG 1022

QY 121 TACACCGTGACCTGAGAGCTTCAAGTGGGAGCTGCGACGCTGATGATGACGCG 240  
DB 1223 TACACCGTGATCTGCTCTTTGAAAGTGGTGAGTGGCTTCCCTATGATGACGCT 1082  
QY 241 CTGCGCTGTGCTCCGCGCAAGTAA 264  
DB 1083 CTGCGCTGTGCTCCGCGCAAGTAA 1106

## RESULT 3

US-09-727-311-5  
Sequence 5, Application US/39727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William E.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
NUMBER OF SEQUENCES: 61  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 200C  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P31 9072  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7294 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-5

Query Match 54.5%; Score 144; DB 9; Length 7294;  
Best Local Similarity 71.6%; Pred. No. 1,76-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGCTGAGATCAAGCCAGCCAGCTGCTACGCTGAGACCTGGCAAGC 50  
DB 843 ATGATTAAGTGAATTAACCATCTCAAGCCCAATTACTCTTCTGTGGTTCT 902  
QY 61 CGCCAGGCGAAGCCCTACAGCCTTAAGCAGAGCTGCTACGCTGAGACCTGGCAAGC 120  
DB 963 CGTCAGGCGAAGCCCTATTACCTGAATGAGCAGCTTTGTACGTTGATTTGGGTAAGAA 962  
QY 121 TACCCGCTGCTGGAAGATCAAGCTGAGAGGCGCAGCCGCTTACGCTGAGCTG 180  
DB 963 TATCCGCTGCTGGAAGATCAAGCTGAGAGGCGCAGCCGCTTACGCTGAGCTG 1022  
QY 121 TACACCGTGACCTGAGAGCTTCAAGTGGGAGCTGCGACGCTGATGATGACGCG 240  
DB 1083 TACACCGTGATCTGCTCTTTGAAAGTGGTGAGTGGCTTCCCTATGATGACGCT 1082  
QY 241 CTGCGCTGTGCTCCGCGCAAGTAA 264

Db 1083 CTGGCCCTGCTCCGGCTAAGTAA 1106

## RESULT 4

US-09-727-311-2  
Sequence 2, Application US/09727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2200  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentia Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION/DOCKET NUMBER: 31,815  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7320 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-2

Query Match 54.5%; Score 144; DB 9; Length 7320;  
Best Local Similarity 71.6%; Pred. No. 1,7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGAGATCAAGCCGACGCGCCAGCTTACGACCCGCGAGCGGCTGAGC 60  
Db 843 ATGATTAAGTTGAATTAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTTCT 902  
QY 61 CGCCAGGCGAAGCCCTACAGCCCTGAGCGAGCGAGCTTGTCTAGCTGAGCTGGGCAAGG 120  
Db 903 CGTACGAGGCAAGCCTATTCACTGAGTGAAGCGAGCTTTGTACGTGATTGGGTAATGAA 962  
QY 121 TACCCCGTGTGAGTGAAGTCAACCTGAGCGAGGCGACGCCCTACGCCGCCCGGCTG 180  
Db 963 TATCCGCTTCTTGTCAAGTAACTCTGTATGAAGGTCAAGCCGCTAAGCGCTGCTG 1022  
QY 181 TACACCGTGCACCTGAGCGAGCTTCAAGGTGCGGCAAGTGGGCAAGCTGATGATGCAACGCC 240  
Db 1023 TACACCGTGCATCTGCTCTTCAAGGTGCTGAGTGGTGGTCCCTATGATGATGACCGT 1082  
QY 241 CTGGCCCTGCTGCGCCGCAAGTAA 264  
Db 1083 CTGGCCCTGCTGCGGCTAAGTAA 1106

RESULT 5

US-09-727-311-6  
Sequence 6, Application US/09727311  
Patent No. US20010024782A1

GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2200  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentia Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/727,311  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/767,436  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A  
REGISTRATION/DOCKET NUMBER: 31,815  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ. ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7394 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-09-727-311-6

Query Match 54.5%; Score 144; DB 9; Length 7394;  
Best Local Similarity 71.6%; Pred. No. 1,7e-29;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGGTGAGATCAAGCCGACGCGCCAGCTTACGACCCGCGAGCGGCTGAGC 60  
Db 843 ATGATTAAGTTGAATTAACCATCTCAAGCCCAATTACTACTGCTTCTGCTTTCT 902  
QY 61 CGCCAGGCGAAGCCCTACAGCCCTGAGCGAGCGAGCTTGTCTAGCTGAGCTGGGCAAGG 120  
Db 903 CGTACGAGGCAAGCCTATTCACTGAGTGAAGCGAGCTTTGTACGTGATTGGGTAATGAA 962  
QY 121 TACCCCGTGTGAGTGAAGTCAACCTGAGCGAGGCGACGCCCTACGCCGCCCGGCTG 180  
Db 963 TATCCGCTTCTTGTCAAGTAACTCTGTATGAAGGTCAAGCCGCTAAGCGCTGCTG 1022  
QY 181 TACACCGTGCACCTGAGCGAGCTTCAAGGTGCGGCAAGTGGGCAAGCTGATGATGCAACGCC 240  
Db 1023 TACACCGTGCATCTGCTCTTCAAGGTGCTGAGTGGTGGTCCCTATGATGATGACCGT 1082  
QY 241 CTGGCCCTGCTGCGCCGCAAGTAA 264  
Db 1083 CTGGCCCTGCTGCGGCTAAGTAA 1106

## RESULT 6

US-09-727-311-4  
Sequence 4, Application US/09727311  
Patent No. US20010024782A1  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.







CURRENT FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204689  
 PRIOR FILING DATE: 2002-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 1  
 LENGTH: 9025608  
 TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (4187715)  
 OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Query Match 20.5%: Score 54.2; DB 14: Length 9025608;  
 Best Local Similarity 54.3%; Pred. No. 2.1e-05;  
 Matches 129; Conservative 0; Mismatches 103; Indels 3; Gaps 1

QY 15 GATCAAGCCGACGCGCCAGTTCACACCCGACGCGAGGTAGAGCCGACGACAGC 74  
 DB 1276169 GATCTACACGACGCGCCGACGAGCGGACGCGCATCTGAGACACCATCCGACACCTC 1276229  
 QY 75 CTACAGCCTGACACGACGAGCTGTGCTACGTGACCTGAGCTGACGAGTACCCGCTGCTGGT 134  
 DB 1276229 CACCTGCTGACGCGCCGCTGAGGACGACGAGGCTAGCGGACCTACATGCTGCTGCT 1276288  
 QY 135 GAAGATACCCCTGACACGAGGACCA---GCCCGCTTACGCGCCCGGCTGTACACCTGTGA 191  
 DB 1276289 GGAGACCGAGATGACCTCGGAGGACGCTGCTGACGACCCGACCCGACCTGCTGCA 1276348  
 QY 192 CCGAGCAGCTTAAAGTGGCGGACGCTTGGGACGCGGTGATGATGAGCGGCTGGCG 246  
 DB 1276349 CGAGTGAACCTGGCGGACGAGACCGCATCGCTTACCGGCTGCTGAGCGGCTGCTC 1276403

RESULT 12  
 US-10-029-386-22785  
 Sequence 22785, Application US/10029386  
 Publication No. US20030194703A:  
 GENERAL INFORMATION:  
 APPLICANT: Penn, Sharon G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 FILE REFERENCE: AROMICA-X-2  
 CURRENT APPLICATION NUMBER: US/10/029,386  
 CURRENT FILING DATE: 2001-12-20  
 NUMBER OF SEQ ID NOS: 34288  
 SOFTWARE: Anomax Sequence Listing Engine Vers. 1.1  
 SEQ ID NO 22785  
 LENGTH: 759  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AC004737.1  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.43  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4  
 OTHER INFORMATION: EST HUMAN HIT: AW69906.1, EVALU 0.00e+00  
 OTHER INFORMATION: SWISSPROT HIT: Q28142, EVALU 3.10e-01  
 OTHER INFORMATION: KT HIT: A006529.1, EVALU 1.30e-08  
 US-10-029-386-22785

Query Match 19.1%: Score 50.4; DB 12: Length 759;  
 Best Local Similarity 51.3%; Pred. No. 0.00022;  
 Matches 117; Conservative 0; Mismatches 111; Indels 0; Gaps 0

QY 34 CAGTTACACCGCGGCGCTGAGCGCGGACGAGCGCTTACAGCTGACGAGCAG 93  
 DB 439 CCGTTCCACACACACAGCTGCGCGGCTGCGCGCAACATGTGCAAGACAAAGTGCTC 498

QY 34 CTGTCTACGTGACACCTGAGGCAAGAGTACCCGCTGTGTGAGATACCTGAGAGAG 153  
 DB 499 AGCGACACCTGCGCGAGCGCGGACGAGCGGAGCTTACAGCGGCTGTCTGACGTG 558  
 QY 154 GGGCAGCGCGGCTTACGCGCGCGGCTTACACCGGTGACCTGAGCAGCTTGMAGTGGG 213  
 DB 559 GAGACGCGCGGACGAGCTTCTGATCCATGAGGCTGCTGAGCGGAGCGTGGCTTTC 618  
 QY 214 CAGTTGCGACACCTGATGATGACCGGCTTGGCTTGGTGGCGGCGAG 261  
 DB 619 CCGCGCGCGGCTTACCCATGACAGCGCTCATTTAGAGAGCGCGGAGAG 666

RESULT 13  
 US-10-259-165-321  
 Sequence 321, Application US/11259165  
 Publication No. US20030135888A1  
 GENERAL INFORMATION:  
 APPLICANT: Zhu, Tong  
 APPLICANT: Wang, Xun  
 APPLICANT: Chang, Hui-song  
 APPLICANT: Briggs, Steven P.  
 APPLICANT: Cooper, Bret  
 APPLICANT: Glazedrock, Jane  
 APPLICANT: Goff, Stephen A.  
 APPLICANT: Katagiri, Fumiyaki  
 APPLICANT: Kieps, Joel  
 APPLICANT: Moughamer, Todd  
 APPLICANT: Privart, Nicholas  
 APPLICANT: Rieke, Darrel  
 TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENC  
 FILE REFERENCE: 72030-NP  
 CURRENT APPLICATION NUMBER: US/10/259,165  
 CURRENT FILING DATE: 2002-09-26  
 PRIOR APPLICATION NUMBER: US 60/370,620  
 PRIOR FILING DATE: 2002-04-04  
 PRIOR APPLICATION NUMBER: US 60/368,387  
 PRIOR FILING DATE: 2002-03-27  
 PRIOR APPLICATION NUMBER: US 60/325,277  
 PRIOR FILING DATE: 2001-09-26  
 NUMBER OF SEQ ID NOS: 782  
 SOFTWARE: PatentList v1 version 3.0.4 (C) 2001 Syngenta  
 SEQ ID NO 321  
 LENGTH: 972  
 TYPE: DNA  
 ORGANISM: Oryza sativa

FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (135)..(135)  
 OTHER INFORMATION: Y = cytosine or thymine  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (139)..(139)  
 OTHER INFORMATION: m = adenine or cytosine  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (156)..(156)  
 OTHER INFORMATION: Y = cytosine or thymine  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (169)..(169)  
 OTHER INFORMATION: X = adenine or guanine  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (197)..(197)  
 OTHER INFORMATION: X = guanine or thymine  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (199)..(199)  
 OTHER INFORMATION: X = adenine or guanine  
 US-10-259-165-321

Query Match 17.7%: Score 46.6; DB 12: Length 972;







```

; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
US-08-440-787A-1

```

Query Match:	54.5%;	Score 144;	DE 1;	Length 7294;
Best Local Similarity	71.6%;	Pred. No. 1.4e-21;		
Matches 189; Conservative	0;	Mismatches 75;	Indels 0;	Gaps 0

QY	ATATCAAGAGTGGAGATCAAGCCGACGAGCCAGTTCACACACCCGACGCGTGAAC	60
DQ	ATGATTAAGTGAATTAACCATCTGACGCCCATTTACTATCTGTTGGTGTTC	902
QY	CGCCAGGGCAAGCCCTACAGCCTGAACAGACACTGTCTACTGCTGACCTGGCAACGAG	120
DQ	CGTCAGGGCAAGCCTATTACCTGATGATGACAGCCTTCTTACGTTGATTTGGGTAATGAA	962
QY	TACCCCGTGTGGTGAAGATCAACCCCTGGACGAGGCCACGCGCTCACTGCGCCCGACCG	180
DQ	TATCCGATTTCTTCAAGATTACCTTTGATGAAGCTCAGCCAGGCCATGCGCCTGGATTG	1022
QY	TACACGCTGACCTTGAGACGCTTCAAGGTCGCGCAGTTCGGACAGCCTGATGATCGACCGC	240
DQ	TACACGCTCATCTGTCCCTCTTCAAAAGTGTGACGTTCGGTCCCTATGATTAAGCCG	1052
QY	CTGGGCGCTGGTGGCCGCCCAATGA	264
DQ	CTGGGCGCTGCTTCGGGCTAAAGTGA	1136

```

1      RESULT 5
2      US-CR-440-787A-5
3      : Sequence 5, Application US/08440787A
4      : Patent No. 570434
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Huse, William J.
9      : TITLE OF INVENTION: Soluble Peptides Having Constrained,
10     : TITLE OF INVENTION: Secondary Conformation in Solution and Method of Making
11     : TITLE OF INVENTION: Same.
12     : NUMBER OF SEQUENCES: 174
13     :
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: Campbell & Flores LLP
16     : STREET: 4370 La Jolla Village Drive, Suite 700
17     : CITY: San Diego
18     : STATE: California
19     : COUNTRY: USA
20     :
21     : ZIP: 92122
22     :
23     : COMPUTER READABLE FORM:
24     : MEDIUM TYPE: Floppy disk
25     : COMPUTER: IBM PC compatible
26     : OPERATING SYSTEM: PC-DOS/MS-DOS
27     : SOFTWARE: Patent'n Release #1.0, Version #1.25
28     :
29     : CURRENT APPLICATION DATA:
30     : APPLICATION NUMBER: US/08/440,787A
31     : FILING DATE: 15-MAY-1995
32     : CLASSIFICATION: 435
33     :
34     : PRIOR APPLICATION DATA:
35     : APPLICATION NUMBER: US 07/978,893
36     : FILING DATE: 13-NOV-1992
37     :
38     : ATTORNEY/AGENT INFORMATION:
39     : NAME: Campbell, Cathryn A.
40     : REGISTRATION NUMBER: 31,815
41     : REFERENCE/DOCKET NUMBER: P-IX 1566
42     :
43     : TELECOMMUNICATION INFORMATION:
44     : TELEPHONE: (619) 535-9001
45     : TELEFAX: (619) 535-8949
46     :
47     : INFORMATION FOR SEQ ID NO: 5:
48     : SEQUENCE CHARACTERISTICS:
49     : LENGTH: 7294 base pairs
50     : TYPE: nucleic acid
51     : STRANDEDNESS: both
52     : TOPOLOGY: circular
53     :
54     : US-08-440-787A-5

```

Query Match:	54.5%;	Score 14;	DB 1;	Length 7294;	
Best Local Similarity	71.6%;	Pred. No. 1.4e-22;			
Matches 189;	Conservative	0;	Mismatches	75;	
		Indels	0;	Gaps	0;

2Y	1	TGATCGAAGGGGAGATCAAGCCCAAGGCGCAAGTTACCAACCCCGAGCGCGGTGAGC	60
2Y	643	ATGATTTAAAGTTGAAATTTAAACAATCTCAAGCCCAATTACTACAGCTTTGTGGTATTC	902
2Y	61	CGCAGAGCGAAGCCCTAGAGCTGAAAGAGCAAGCTGTCTAGTGGACCTGACACACAG	120
2Y	933	CGTCAGGCGAAGCCCTATTCACTGAAATGAGCAGACCTTTGAGGTGATTTGGATATATGAA	962
2Y	121	TACCCCGTGGTGGAGATCACTCGAGCGAGAGCGTCAAGCCCGCCTCAAGCCCGCGCTGG	150
2Y	965	TTTCCGGTCTTGTCAAGATTACTCTTTGAGGAAGGTCAAGCTCAAGCTTACGAGTGGCTTG	1022
2Y	181	TACACGTGCACCTGAGCAAGCTTCAAGTGTGGCCATTTGGCAAGCTTGATATGATACCG	240
2Y	1023	TACACGCTTCATCTGTCTCTTTCGAAAGTGGTCAGTTCGATTCCTTATGATATGACCGT	1082
2Y	241	CTGGGCGTGGTCCCGCGCAAGTAA	264
2Y	1083	CTGGGCGCTCTTCGCGTAAAGTAA	1106

```

1      RESULT 6
2      US-08-367-685-1
3      Sequence 1, Application US/08367685
4      Patent No. 6236930
5      GENERAL INFORMATION:
6      APPLICANT: Huse, William D.
7      TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF
8      TITLE OF INVENTION: RANDOMIZED PEPTIDES
9      NUMBER OF SEQUENCES: 6;
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
12     STREET: 444 South Flower Street, Suite 2000
13     CITY: Los Angeles
14     STATE: California
15     COUNTRY: United States
16     ZIP: 90071
17
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: PatentIn Release #1.0, Version #1.25
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/367,685
25
26     FILING DATE:
27     CLASSIFICATION: 435
28     PRIOR APPLICATION DATA:
29     APPLICATION NUMBER: US/08/110,494
30     FILING DATE:
31     APPLICATION NUMBER: US/07/767,436
32     FILING DATE:
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Campbell, Cathryn A
35     REGISTRATION NUMBER: 31,615
36     REFERENCE/DOCKET NUMBER: P11 9072
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE: (619) 535-9001
39     TELEFAX: (619) 535-8849
40
41     INFORMATION FOR SEQ ID NO. 1:
42     SEQUENCE CHARACTERISTICS:
43     LENGTH: 7294 base pairs
44     TYPE: nucleic acid
45     STRANDEDNESS: both
46     TOPOLOGY: circular
47
48     US-08-367-685-1
49
50     Query Match      54.5%   Score 144;   DB 3;   length 7294;
51     Best Local Similarity  71.6%   Pred. No. 1,4e-21;

```

Matches 189: Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACCCGAGCCGCTGAGC 60

DB 843 ATGATTAAAGTTAAATTAAACCATCTCAAGCCCAATTACTACTGTTTCTGTTCT 902

QY 61 CGCCAGGCGAAGCCCTTACAGCTTGAAGAGAGCTGCTGAGCTGAGCTGAGCTGAG 120

DB 903 CGTCAGGCGAAGCCCTTATTCAGTGAATGAGCAGCTTGTGAGCTTATTTGGATATGA 962

QY 121 TACCCGCTGCTGAGTGAAGATCAACCTGAGAGAGCTGAGCCGCTGAGCCGCTG 180

DB 963 TATCCGCTTCTTCAAGATTAATCTGATGAGAGTCAAGCCAGCCGCTGAGCTG 1022

QY 181 TACACGCTGACCTGAGAGCTTCAAGTGGCGAGCTTGGGAGCTGATGATGAGCCG 240

DB 1023 TACACGCTTATCTGCTCTTTCATGAAGTGGTCAAGCTTCCCTTATGATGAGCCG 1082

QY 241 CTGCGCTGCTGAGCCGCGAGTAA 264

DB 1083 CTGCGCTGCTGAGCCGCGAGTAA 1106

## RESULT 7

US-08-367-685-5

Sequence 5, Application US/08367685

Patent No. 6258530

GENERAL INFORMATION:

APPLICANT: Huse, William D.

TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF

TITLE OF INVENTION: RANDOMIZED PEPTIDES

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder, Rueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: California

COUNTRY: United States

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version 4.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/367,685

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/110,494

FILING DATE:

APPLICATION NUMBER: US/07/767,436

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P31 9072

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 7294 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: circular

US-08-367-685-5

## Query Match

54.5%; Score 144; DB 3; Length 7294;

Best Local Similarity 71.6%; Pred. No. 1.4e-21;

Matches 189: Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACCCGAGCCGCTGAGC 60

DB 843 ATGATTAAAGTTAAATTAAACCATCTCAAGCCCAATTACTACTGTTTCTGTTCT 902

QY 61 CGCCAGGCGAAGCCCTTACAGCTTGAAGAGAGCTGCTGAGCTGAGCTGAGCTGAG 120

DB 903 CGTCAGGCGAAGCCCTTATTCAGTGAATGAGCAGCTTGTGAGCTTATTTGGATATGA 962

QY 121 TACCCGCTGCTGAGTGAAGATCAACCTGAGAGAGCTGAGCCGCTGAGCCGCTG 180

DB 963 TATCCGCTTCTTCAAGATTAATCTGATGAGAGTCAAGCCAGCCGCTGAGCTG 1022

QY 181 TACACGCTGACCTGAGAGCTTCAAGTGGCGAGCTTGGGAGCTGATGATGAGCCG 240

DB 1023 TACACGCTTATCTGCTCTTTCATGAAGTGGTCAAGCTTCCCTTATGATGAGCCG 1082

QY 241 CTGCGCTGCTGAGCCGCGAGTAA 264

DB 1083 CTGCGCTGCTGAGCCGCGAGTAA 1106

## RESULT 8

PCT-US91-07141-2

Sequence 1, Application PC/US9107141

GENERAL INFORMATION:

APPLICANT: Huse, William D.

TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF

TITLE OF INVENTION: RANDOMIZED PEPTIDES

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder, Rueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: California

COUNTRY: United States

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version 4.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/07141

FILING DATE: 19910927

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P31 9072

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7294 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: both

TOPOLOGY: circular

PCT-US91-07141-2

## Query Match

54.5%; Score 144; DB 5; Length 7294;

Best Local Similarity 71.6%; Pred. No. 1.4e-21;

Matches 189: Conservative 0; Mismatches 75; Indels 0; Gaps 0

QY 1 ATATCAAGGTGAGATCAAGCCAGCCAGCCAGCTTACACCCGAGCCGCTGAGC 60

DB 843 ATGATTAAAGTTAAATTAAACCATCTCAAGCCCAATTACTACTGTTTCTGTTCT 902

QY 61 CGCCAGGCGAAGCCCTTACAGCTTGAAGAGAGCTGCTGAGCTGAGCTGAGCTGAG 120

DB 903 CGTCAGGCGAAGCCCTTATTCAGTGAATGAGCAGCTTGTGAGCTTATTTGGATATGA 962

QY 121 TACCCGCTGCTGAGTGAAGATCAACCTGAGAGAGCTGAGCCGCTGAGCCGCTG 180

DB 963 TATCGGCTTCTGTGTAAGATTAAGTTCAGAGGTCAGGCTATGCGGCTGCTG 1022

QY 181 TACACGCTGACCTGAGAGGTTCAAGGTCGCGGAGTTGCGGAGGCTGATGATGACGCG 240

DB 1023 TACACGCTGATCTGCTCTCTTCAAAAGTTGCTAGAGTTGGTTCCCTTATGATGACGCT 1082

QY 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264

DB 1083 CTGCGGCTGCTGCGGCTGAGTAA 1106

## RESULT 9

PCT-US91-07141-5

Sequence 5, Application PC/TUS9107141

GENERAL INFORMATION:

APPLICANT: Huse, William D.

TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF

TITLE OF INVENTION: RANDOMIZED PEPTIDES

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESS: Pretty, Schroeder, Brueggeman & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: California

COUNTRY: United States

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/07141

FILING DATE: 19910927

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P31 9072

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-535-9001

TELEFAX: 619-535-8949

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 7294 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: both

TOPOLOGY: circular

PCT-US91-07141-5

Query Match 54.5%; Score 144; DB 5; Length 7294;

Best Local Similarity 71.6%; Pred. No. 14e-21;

Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCGACGCGGCAATTCACACCGGCGGCGCTGAC 60

DB 843 ATGATTAAGTTAAATTAACATCTCAAGCCCAATTAATGATGTTCTGTTCT 902

QY 61 CGCCAGGCGAAGGCTTACAGGCTGAGAGGAGGAGGCTGCTAGCTGAGGCTGAGGAG 120

DB 903 CATTAGGGGAGAGGCTTATTCATGATGAGGAGGCTTGTACGTTGATTGGGTAATGAA 962

QY 121 TACCCGCTGCTGTGAGATCAAGCTTGAAGAGGCGGCGGCTTACGCGGCGGCTG 180

DB 963 TATCCGCTTCTGTCAAGATTAATCTTGTGATGAAAGTCAAGCCGATATCGGCTG 1022

QY 181 TACACGCTGACCTGAGAGGTTCAAGGTCGCGGAGTTGCGGAGGCTGATGATGACGCG 240

DB 1023 TACACGCTGATCTGCTCTCTTCAAAAGTTGCTAGAGTTGGTTCCCTTATGATGACGCT 1082

QY 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264

DB 1083 CTGCGGCTGCTGCGGCTGAGTAA 1106

## RESULT 10

US-08-464-136-2

Sequence 2, Application US/08464136

Patent No. 5698426

GENERAL INFORMATION:

APPLICANT: HUSE, WILLIAM D.

TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF

TITLE OF INVENTION: HETEROMERIC RECEPTORS

NUMBER OF SEQUENCES: 75

CORRESPONDENCE ADDRESS:

ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMAN & CLARK

STREET: 444 S. FLOWER STREET, SUITE 200

CITY: LOS ANGELES

STATE: CALIFORNIA

COUNTRY: UNITED STATES

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/464-136

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATHRYN A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P31 8882

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-535-9001

TELEFAX: 619-535-8949

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 7317 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: circular

US-08-464-136-2

Query Match 54.5%; Score 144; DB 1; Length 7317;

Best Local Similarity 71.6%; Pred. No. 1.4e-21;

Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCGACGCGGCAATTCACACCGGCGGCGCTGAC 60

DB 843 ATGATTAAGTTAAATTAACATCTCAAGCCCAATTAATGATGTTCTGTTCT 902

QY 61 CGCCAGGCGAAGGCTTACAGGCTGAGAGGAGGAGGCTGCTAGCTGAGGCTGAGGAG 120

DB 903 CGTCAGGCGAAGGCTTATTCATGATGAGGAGGCTTGTACGTTGATTGGGTAATGAA 962

QY 121 TACCCGCTGCTGTGAGATCAAGCTTGAAGAGGCGGCGGCTTACGCGGCGGCTG 180

DB 963 TATCCGCTTCTGTCAAGATTAATCTTGTGATGAAAGTCAAGCCGATATCGGCTG 1022

QY 181 TACACGCTGACCTGAGAGGTTCAAGGTCGCGGAGTTGCGGAGGCTGATGATGACGCG 240

DB 1023 TACACGCTGATCTGCTCTCTTCAAAAGTTGCTAGAGTTGGTTCCCTTATGATGACGCT 1082

QY 241 CTGCGGCTGCTGCGGCGGCAAGTAA 264

DB 1083 CTGCGGCTGCTGCGGCTGAGTAA 1106

## RESULT 11

US-08-349-131-2

Sequence 2, Application US/08349131;

Patent No. 5871974

GENERAL INFORMATION:

APPLICANT: HUSE, WILLIAM D.





COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/07149  
FILING DATE: 19910927  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CAMPBELL, CATHRYN A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-IX 1586  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-535-9001  
TELEFAX: 619-535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7317 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: both  
TOPOLOGY: circular  
PCT-US91-07149-2

Query Match 54.5%; Score 144; DB 5; Length 7317;  
Best Local Similarity 71.6%; Pred. No. 1,4e-21;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTACCAACCCGAGCCGCTGAGC 60  
DB 843 ATGATCAAGTGAATTAACATCTCAAGCCCAATTACACACCTTCTGCTTCT 902  
QY 61 CGCCAGGGCAAGCCCTACAGCCTGAAGAGAGAGCTGTCTAGCTGAGCTGAGCAACAG 120  
DB 903 CGTACAGGGCAAGCCTTATTCATCAATGAGAGAGCTGTGATGATTTGAGTATGAA 962  
QY 121 TACCCGCTGTGTGAAGTCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 963 TATCCGCTGTGTGAAGTCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022  
QY 181 TACACCGTGTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 1023 TACACCGTGTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082  
QY 241 CTGCGCCTGTGTGAG 264  
DB 1083 CTGCGCCTGTGTGAG 1106

## RESULT 14

US-08-440-787A-2  
Sequence 2, Application US/08440787A  
Patent No. 5770434  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: Soluble Peptides Having Constrained,  
TITLE OF INVENTION: Secondary Conformation in Solution and Method of Making  
NUMBER OF SEQUENCES: 174  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,787A  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435

Prior Application Data:  
APPLICATION NUMBER: US 07/378,593  
FILING DATE: 12-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-IX 1586  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7320 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
US-08-440-787A-2

Query Match 54.5%; Score 144; DB 1; Length 7320;  
Best Local Similarity 71.6%; Pred. No. 1,4e-21;  
Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1 ATGATCAAGTGGAGATCAAGCCAGCCAGCCAGCTTACCAACCCGAGCCGCTGAGC 60  
DB 843 ATGATCAAGTGAATTAACATCTCAAGCCCAATTACACACCTTCTGCTTCT 902  
QY 61 CGCCAGGGCAAGCCCTACAGCCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 903 CGTACAGGGCAAGCCTTATTCATCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
QY 121 TACCCGCTGTGTGAAGTCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 963 TATCCGCTGTGTGAAGTCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022  
QY 181 TACACCGTGTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 1023 TACACCGTGTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082  
QY 241 CTGCGCCTGTGTGAG 264  
DB 1083 CTGCGCCTGTGTGAG 1106

## RESULT 15

US-08-367-693-2  
Sequence 2, Application US/08367693  
Patent No. 6285530  
GENERAL INFORMATION:  
APPLICANT: Huse, William D.  
TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF  
TITLE OF INVENTION: RANDOMIZED PEPTIDES  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Prety, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/367,693  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,454  
FILING DATE:  
APPLICATION NUMBER: US/07/767,436  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P31 9072  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9000  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7323 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: circular  
 US-08-367-685-2

Query Match 54.5%; Score 144; DB 3; Length 7323;  
 Best Local Similarity 71.6%; Pred. No. 1,4e-21;  
 Matches 189; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY	1	ATGATCAAGGTGGAGATCAAGCCAGCCAGCCCAAGTTCACCAACCGGACGGGCGTGAAC	60
DB	843	ATGATTAAGTTGAAATTAACCACTCAAGCCCAATTACACTGCTTCGGTGTCTTCT	902
QY	61	CGCCAGGCGCAAGCCCTACAGCTTAACGAGCAAGCTGCTACCTGGACCTGGCAACGAG	120
DB	903	CGTCAGGCGCAAGCCCTATTCACGATGAGACAGCTTGTACGTTGATTTGGGTATGAA	962
QY	121	TACCCCGTGGTGGAGATCAACCTGGACGAGGCGCCAGCCCGCCCGGCGCTG	180
DB	963	TATCGGTTCTTGCAAGATTACTTTGATGAAGGTCAGCCAGCTATGGGCTGGTCTG	1022
QY	181	TACACCGTGACCTGACAGCTCAAGCTCAAGCTGGCCAGTTGGGACGCTGATGATGACCGC	240
DB	1023	TACACCGTTCATGCTGCTCTCTTCAAGCTGGTCAAGTGGTGGTTCCTTATGATGACCGT	1082
QY	241	CTGGGCGTGGTGGCCGCAAGTAA	264
DB	1083	CTGGGCGTGGTGGGCTAAGTAA	1106

Search completed: November 7, 2003, 12:56:58  
 Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 13:19:21 Search time 1510 seconds

(without alignments) 7152.412 Million cell updates/sec

Title: US-09-622-500b-3

Perfect score: 264

Sequence: 1 atgacaaagtgagagatca.....gacctggtgcccgaagaa 264

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database:

GenBdb:  
1: gb\_da:  
2: gb\_hg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pn:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vl:  
15: em\_da:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_or:  
22: em\_ov:  
23: em\_pac:  
24: em\_ph:  
25: em\_pl:  
26: em\_ro:  
27: em\_sts:  
28: em\_un:  
29: em\_vl:  
30: em\_hg\_hum:  
31: em\_hg\_iny:  
32: em\_hg\_other:  
33: em\_hg\_mus:  
34: em\_hg\_pln:  
35: em\_hg\_rdd:  
36: em\_hg\_mam:  
37: em\_hg\_vrt:  
38: em\_sy:  
39: em\_hgo\_hum:  
40: em\_hgo\_mus:  
41: em\_hgo\_other:

Pred. No. is the number of results predicted by chance to have a

Score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	FR	ID	Description
1	21	9.0	230905	2	AC132180	AC132180 Rattus no
2	20	7.6	6774	9	HSB136570	AC132180 Homo sapi
3	20	7.6	15124	2	AC108763	AC108763 Oryza sat
4	20	7.6	347950	1	AF003113	AF003113 Mesocric
5	19	7.2	3624	3	AF132169	AF132169 Drosophi
6	19	7.2	3773	1	PSENGRC	PSENGRC Pseudomora
7	13	7.2	10711	1	AE004489	AE004489 Pseudomon
8	13	7.2	10024	1	AE012081	AE012081 Xanthomon
9	13	7.2	11854	1	AE012540	AE012540 Xanthomon
10	19	7.2	14616	9	HS064672	HS064672 Human DNA
11	19	7.2	20686	2	AC017508	AC017508 Drosophi
12	19	7.2	74433	2	AC015098	AC015098 Drosophi
13	19	7.2	107056	9	AL596822	AL596822 Human CNA
14	13	7.2	135055	2	AC010556	AC010556 Drosophi
15	13	7.2	135055	2	AC011909	AC011909 Drosophi
16	19	7.2	144305	2	AC115056	AC115056 Mus muscu
17	19	7.2	150959	2	AC019250	AC019250 Homo sapi
18	19	7.2	16617	2	AC126624	AC126624 Mus muscu
19	19	7.2	166185	10	AL928797	AL928797 Mouse DNA
20	19	7.2	168514	3	AC023350	AC023350 Drosophi
21	19	7.2	174920	3	AC015701	AC015701 Drosophi
22	19	7.2	177205	2	AC075217	AC075217 Mus muscu
23	19	7.2	177205	2	AC075217	AC075217 Mus muscu
24	19	7.2	188225	2	AC108587	AC108587 Rattus no
25	19	7.2	189461	2	AC118247	AC118247 Mus muscu
26	19	7.2	190598	10	AC038892	AC038892 Mus muscu
27	19	7.2	202405	2	AC124556	AC124556 Mus muscu
28	19	7.2	203261	2	AC011374	AC011374 Homo sapi
29	19	7.2	225031	2	AC131174	AC131174 Mus muscu
30	19	7.2	225073	2	AC132397	AC132397 Mus muscu
31	19	7.2	225275	2	AC111406	AC111406 Rattus no
32	19	7.2	235936	2	AC036473	AC036473 Rattus no
33	19	7.2	242510	2	AC037243	AC037243 Rattus no
34	19	7.2	245792	2	AC138101	AC138101 Rattus no
35	19	7.2	269777	2	AC037564	AC037564 Rattus no
36	19	7.2	266511	2	AC133768	AC133768 Rattus no
37	19	7.2	273980	2	AC111446	AC111446 Rattus no
38	19	7.2	277331	2	AC110151	AC110151 Rattus no
39	19	7.2	294272	3	AE033595	AE033595 Drosophi
40	19	7.2	298700	1	AP035951	AP035951 Bradyrhiz
41	19	7.2	305469	3	AE035601	AE035601 Drosophi
42	19	7.2	305550	1	AP005030	AP005030 Streptomy
43	18	6.8	466	8	AY063045	AY063045 Arabidops
44	18	6.8	587	6	BD155522	BD155522 Primer fo
45	18	6.8	646	8	AY034898	AY034898 Arabidops

# ALIGNMENTS

RESULT:  
AC132180  
LOCUS: Rattus norvegicus clone CH23C.123H15, WORKING DRAFT SEQUENCE, 3  
DEFINITION: Rattus norvegicus clone CH23C.123H15, WORKING DRAFT SEQUENCE, 3  
ACCESSION: AC132180  
VERSION: AC132180.3 G1:23664858  
KEYWORDS: HTG, HTGS, PHASE1, HTGS, DRAFT, HTGS, FULLTOP.  
SOURCE: Rattus norvegicus (Norway rat)  
ORGANISM: Rattus norvegicus  
Buxaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,  
Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae,  
Rattus.  
REFERENCE: 1 (bases 1 to 230905)  
Muzny,D,Marie, Metker,M, Lee, Abramson,S, Adams,C, Alder,J.,

Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Balcwin, D., Bandaranaike, D., Barber, M., Barrstead, M., Behar, M., Biswas, K., Blair, J., B. Ankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dirth, H., DiVya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, C., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, M., Garza, M., Gregorogors, E., Geer, K., Giller, R., Grady, M., Guerra, M., Guevara, W., Guarnate, P., Hagard, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., H. Adun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebirt, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khar, Z., King, J., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Lohsiedel, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathney, S., McLeod, V.P., McNeill, T.Z., Meenen, E., Micosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Narkvis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeme, O., Okwucu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, X., Ret, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Soresle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Wa-dron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

# TITLE JOURNAL

REFERENCE  
Rat Genome Sequencing Consortium.  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Submitted (10-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Oct 10, 2002 this sequence version replaced gi:125397366.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atl/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project information  
Center project name: GNPM  
Center clone name: CH230-123H15  
Summary Statistics  
Assembly program: Phrap, version 0.990329  
Consensus quality: 21358 bases at least Q40  
Consensus quality: 21657 bases at least Q30  
Estimated insert size: 21751 bases at least Q30  
Estimated insert size: 219413; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 225352: contig of 225352 bp in length  
225353 225452: gap of unknown length  
225453 229195: contig of 3743 bp in length  
229196 229295: gap of unknown length  
229296 230905: contig of 1610 bp in length

## FEATURES Source

/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-123H15"  
/note="wgs contig"  
/note="wgs contig"

## misc\_feature

## misc\_feature

BASE COUNT 60473 a 47275 c 47001 g 64040 t 12116 others  
ORIGIN

Query Match 8.0%; Score 21; DB 2; Length 230905;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 AAGCCAGCCAGCCAGTC 39  
DE 50267 AAGCCAGCCAGCCAGTC 5287

## RESULT 2

HSA336570:G  
LOCUS HSA336570 674 bp DNA linear PRI 18-JUL-2002  
DEFINITION Homo sapiens genomic sequence surrounding Not1 site, clone NRI-P124R.  
ACCESSION AC336570  
VERSION AC336570.1 GI:15880988  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE  
AUTHORS  
Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M., Podsenko, R.M., Matushkin, Y.G., Gyancharani, A., Kuravenko, O.V., Levitsky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I., Kisselev, L.L., Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.  
TITLE  
Not1 flanking sequences: a tool for gene discovery and verification

```

of the human genome
Nucleic Acids Res. 30 (14), 3163-3170 (2002).
MEDLINE
1213767
PUBMED
12136098
REFERENCE
2 (bases 1 to 674)
AUTHORS
Zabarovskiy, B.R.
DIRECT SUBMISSION
JOURNAL
Submitted (16-MAY-2003) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag. 3, Box 280, Stockholm S-171 77,
Sweden
FEATURES
Location/Qualifiers
source
1..674
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NR1-P1248"
BASE COUNT
125 a 204 c 226 g 119 t
ORIGIN
Query Match 7.6%; Score 20; DB 9; Length 674;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 162 CGCTACGCGCCCGGCGCTGT 181
|||||
Db 50 CGCTACGCGCCCGGCGCTGT 31

RESULT 3
AC008763 155124 bp DNA linear HTG 31-JAN-2002
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) chromosome 9 clone
OSJNB0004A05, *** SEQUENCING IN PROGRESS ***, 15 ordered pieces.
ACCESSION
AC008763
VERSION
1 GI:18443968
KEYWORDS
HTG; HTGS; PHASE2.
SOURCE
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Cryzeae; Oryza.
REFERENCE
1 (bases 1 to 155124)
Yoon, J.-H., Hahn, J.-H., Yun, D.-W., Lee, J.-S., Lee, M.-C., Eun, N.-Y.
and Kim, H.-I.
Oryza sativa BAC OSJNB5004A05 genomic sequence
TITLE
Oryza sativa BAC OSJNB5004A05 genomic sequence
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 155124)
Hahn, J.-H. and Kim, H.-I.
Direct Submission
JOURNAL
Submitted (31-JAN-2002) Rice Genome Sequencing Project, National
Institute of Agricultural Science and Technology (NIAS7), RDA, 249
Seodun-dong, Suwon 441-707, Korea (E-mail: jhahn@da.go.kr,
Tel: 82-31-290-0309, Fax: 82-31-290-0308)
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 7850; contig of 7850 bp in length
* 7851 7950; gap of unknown length
* 7951 17664; contig of 9714 bp in length
* 17665 17665; gap of unknown length
* 17665 39434; contig of 21670 bp in length
* 39435 39535; gap of unknown length
* 39535 44003; contig of 4469 bp in length
* 44004 44103; gap of unknown length
* 44104 55602; contig of 11499 bp in length
* 55603 55703; gap of unknown length
* 55703 107219; contig of 51517 bp in length

```

```

* 107220 107319; gap of unknown length
* 107320 114663; contig of 7344 bp in length
* 114664 114763; gap of unknown length
* 114764 115098; contig of 2335 bp in length
* 115099 117198; gap of unknown length
* 117199 131565; contig of 14767 bp in length
* 131566 13265; gap of unknown length
* 13266 13414; contig of 2249 bp in length
* 13415 13627; gap of unknown length
* 13628 13913; contig of 4593 bp in length
* 13914 145693; contig of 10492 bp in length
* 145694 149700; gap of unknown length
* 149701 151781; contig of 2082 bp in length
* 151782 151881; gap of unknown length
* 151882 153161; contig of 1280 bp in length
* 153162 153261; gap of unknown length
* 153262 155124; contig of 1863 bp in length.
FEATURES
Location/Qualifiers
source
1..155124
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/variety="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="OSJNB0004A05"
BASE COUNT
43170 a 33383 c 33969 g 43152 t 1450 others
ORIGIN
Query Match 7.6%; Score 20; DB 2; Length 155124;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 194 TGAGCAGCTTCAGGTCGGC 213
|||||
Db 81541 TGAGCAGCTTCAGGTCGGC 81560

RESULT 4
AC003013 347950 bp DNA linear BCT 15-MAY-2002
LOCUS
DEFINITION
Mesorhizobium loti: DNA, complete genome, section 20/21.
ACCESSION
AP003013 SA000022
VERSION
AP003013.2 GI:14027324
KEYWORDS
Mesorhizobium loti
Mesorhizobium loti
ORGANISM
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Idegawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, Y., Matsumoto, Y., Matsuno, A.,
Yoshizuki, Y., Nakayama, S., Nakazaki, N., Saito, S., Sugimoto, M.,
Takeuchi, C., Yada, K. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
MEDLINE
21282930
PUBMED
2124968
REFERENCE
2 (bases 1 to 347950)
Kaneko, T.
Direct Submission
JOURNAL
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yara
1532-3, Kasarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/rhizobase/,
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
COMMENT
On May 11, 2001 this sequence version replaced gi:11994388.
FEATURES
Location/Qualifiers
source
1..347950
/organism="Mesorhizobium loti"

```

```

/moi_type="genomic DNA"
/strain="NAF030399"
/db_xref="taxon:381"
complement(78..1037)
/gene="ml17923"
complement(78..1037)
/gene="ml17923"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB3594.1"
/db_xref="GI:14027325"
/transl_table="MELISQVNLNSLKIVESAASHGCHNTRAGCEOFITASAVSCWKS
:EQCLAPKIQRGNGNVSUTETSETVSVREALERIVAASMEATQSGEHVKKGLV
PFAAAMFPJPEFGRYCDIIMRVSTSYATHEFTSFDLSIRYGGSHFNGSPSDJ
LFREDTFCVSRKLFHEVAGDKVPKVEDCLHFTLLHSDCTQNMQMLGFAGASF
VLSKTSYFDSQVMXYEANAAGMGAFAVANKRAYMADIRAEKLVAFFAVHHPNSAGWY
FVSPHKCLARKVLLPKQWVAEAAITQQLDSEIRDLATEAV"
1150..1986
/gene="ml17924"
1150..1986
/gene="ml17924"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3595.1"
/db_xref="GI:14027326"
/transl_table="XNAHTTPELRYAVNSREAI:GHTAWKVSVFGVAQYLVHYDPAJJ
AAIEFAA:KLKDSHAKH:JDLFTITGAPRF:PEIKELHDKLRJERLDMMGTKLEP
YPLISVGTSTFPMNPRGAVWCDGVPTTELPLAISNPLVGGHLEIVCDSEYGRV
LDSGSEIPNNRMDKHNMYATLGQFVJHRTAP:QVGEVPTVLNORVAKRY
DQNMFLAANDNDHREWNELAEQVW:NC:PAYREFABHPTPAFASSETQVSGNARG
SK"
2032..2958
/gene="ml17925"
2032..2958
/gene="ml17925"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3596.1"
/db_xref="GI:14027327"
/transl_table="MVQHRDPRAGRLFRHGSSISALVFAAGAVALKSTNA:LVKSLI
:ATPVSQVQLQAGALNFALIRPMSSEAPSPVAPRAVAALAVGII:GLVGMV:CY
IAPASVIAENL:VAYTWPLMVAAYAFDNPAPRAL:GLAAV:GFVGHVALISGSR
NMFQGLVGFVAFAGALCMATFVNVGRLATSPKLLDPSGSGVGVALLALWSAGEG
VAMPAGMDLAGLVLAGPVGILGYFWSRALKLEGSGKVAWVAYLTP:ASTGLLTUSG
EQLTTTA:AGAVLV:IGSCIAVGLRSPAQDYV"
2951..3154
/gene="msr7926"
2951..3154
/gene="msr7926"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3597.1"
/db_xref="GI:14027328"
/transl_table="MPCDNERLARQEAHWLIKFEAGAPAYAAKMAKALEQCDKDFEC
ARWRRIE:LDARKIT:SAASK?"
3307..3759
/gene="ml17927"
3307..3759
/gene="ml17927"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB3598.1"
/db_xref="GI:14027329"
/transl_table="XKLSNTEQTV:ELTADIVSYVGNPLPASGLPLIASVSAS
VKLAGAVVESP:LIVANPKSVPEDYICLEDGKKFKSLRHLATCYG:GCDYR
AKWGLPDPYFVAVNYNATRSALAKSTGLGRKKPAAAPAAVAKKAKA"
3916..4065
/gene="msr7929"

```

---

```

3916..4065
/gene="msr7929"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3599.1"
/db_xref="GI:14027330"
/transl_table="MNCNAYFEAKGRSALVLQXVFTLFUELLGGSKPLDRNHCHRA
SLRPR"
complement(4476..4730)
/gene="msl7931"
complement(4476..4730)
/gene="msl7931"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB3600.1"
/db_xref="GI:14027331"
/transl_table="MRSERSLMIQYAAALJHRTSTLRKALPASWPDVRVKTGGSGSALRH
QPGGKWRPQDFCCVADRKTAGTDIEDTGLRNLYFYSC"
4699..4699
/gene="ml17932"
4699..4699
/gene="ml17932"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding component"
/protein_id="BAB3601.1"
/db_xref="GI:14027332"
/transl_table="MLIINDLSRYAGRLJLCHASLTDFPAGTKAGLVGRNQTGTLF
KAT:GDSEFSG:SLRKT:RIQOVAQAP:TEPLIEI:VLKAOVET:ALIEBKAT
DPRIA:DHMLAD:DAHSAESBA:LAG:GFDDAQGPASSFSGWRMRVALAAV
LSEFD:LLDSEFNY:JLEGT:WLENYVSKY:PHTVL:SHDRLL:KRNANSI:VHLDQ
KLT:FWRGYDQFEQY:TEQKEJGKGRVQEAARKHSEFVERFRAKAKARQQR
IKALEMKPIAAVNDSPFSEFPYKIVASPIVALNNVNYTEGOPIKKTLLRI
DADRIALLGANGKSTFAKLLSRKQESQTMVAPGLKVAIFAQHLDLRPEEN
AYEHVRR:MPPEASQKSVGRVAQFGL:TEKMT:FAKOLSGEKARL:WGLSAFSGPN
FILDEPTNHLIDQRESLIHALNEFPQAVILI:SHDRHLLEATARLW:VKDGA:NPV
GDLECYKTLVTGVSQDGRGKREAKASKASPRDAAARAFEFELAKEIRATEALMDR
IKRR:DGIEDELANFAIYEKDFSTATLAKERSQLAQTJAGHEEKWJLMSABYSEGTA
E"
6667..8178
/gene="ml17933"
6667..8178
/gene="ml17933"
/notes="similar to formate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB3602.1"
/db_xref="GI:14027333"
/transl_table="MNOHAKLIRIGHSACPHDPCSTCALEVELLDNRLIGRYVEGAKANT
YTSGVVCAKVARVADRVHHPDLRLKPLIRAGAKGDGVQCBASWZALDLVAEK:IAAE
EKYQSETVMFYFYAASC:AAVTGQVGGGAFHNSGIFKQNEVLGSTRMDDPSIR
YLDHSH:GPV:GASDALYGGPPVTAML:QNTNPNVAPQQLKQGLRDLQFACVH
EQFMTD:ARLADVLVLPATYFLEHDDVYKGGNQHIT:GPKLI:EPPEGRTNHFVIEQL
ARLQVADRPGFGLTEQOHIDIILKRG:LSGFSDLKEOKWVDLOPCEFAHFIKGFHG
SDGKPRPADWTGGAANRPKPSXGLFQGPVRLPEFPDHWZLIEVADDAHPRP:JTSP
ARNFLNST:FASTFVSKKEGSPALLLHPDDADLGLADGDRVEYQNGDGLVLAKEFF
DG:KGVV:ASG:WPNSSHERGEGINVLITGACAFYGGAAFHONKVVLRKAG"
complement(8203..9390)
/gene="ml17934"
complement(8203..9390)
/gene="ml17934"
/notes="similar to dimethylamine monooxygenase"
/codon_start=1
/transl_table=11
/protein_id="BAB3603.1"
/db_xref="GI:14027334"
/transl_table="WELTFSGRE:SRREJEDTTTVALIGAGPAGLAAAC:RCAGVD
PIT:EKEOQAPAMRRHVERVHLHTTKYSSLPFPFPKYPYRVYFRALFYCYL:DAYA
GRFLRPOFGETKAVTQCGWGRVDAASGFLRAKDVVIAAGSYNAEPURPAPAGICTF
TKTGLSHADYRNKAF:FAQGSYLV:MGNT:GAELALDAENGAQFTISVEGQHVIPRE
LPGVFIQWGVMAARLGPQRLNDALFFVILDLVLGRLDKYG:KRPQGG:LEQIVASR:

```

PVIDVGTGKIREGAIKVAPDITEISQGRFADGKHGEFEDAIIFATGYRPGVARELE  
PGIQDRSGVTPKASDLGLYLGHPNAVTLGRIEIGIEAQAIADDSHRLNRKAAE1  
Lpv"

gene complement (9426, .9905)

/gene="ml17935"

CDS complement (9426, .9905)

/gene="ml17935"

/codon\_start=1

/transl\_table=1

/product="transcriptional regulator"

/protein\_id="BAB5184.1"

/db\_xref="GI:14027335"

/translation="VPESSPPDLSFDLAILAVLQRTTQRILIGAVNLAPAVRR

KRMEQAGV:ASNAV:EPAAVGKPI:TFVEVELESERTLIDAARQFQSTPEVQC

YVVTGEADFILVITVADMGAVEALTRKLFSSNVKRFETVAVDRVKVGLTVPD"

10133, .11218

gene /gene="ml17936"

CDS /gene="ml17936"

/note="hypothetical protein"

/codon\_start=1

Query Match 7.6%; Score 20; DB 1; Length 347950;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 20; Conservative C; Mismatches 0; Indels 0; Gaps C;

QY 201 CTTCAAGGTGGCCAGTTCG 220

|||||:|||||:|||||

Db 21:819 CTTCAAGGTGGCCAGTTCG 211838

# RESULT 5

AF132169

LOCUS

DEFINITION Drosophila melanogaster clone Lb21405 unknown mRNA.

ACCESSION

AF132169.1

VERSION

AF132169.1

KEYWORDS

3' UTR, 5' UTR, 3' UTR

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Ekcoperlygota; Diptera; Brachycera; Muscomorpha;

Spheroidea; Drosophilidae; Drosophila.

1 (bases 1 to 3624)

Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,

Tsang,G., Agbayani,A., Arcana,T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,

Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,

Li,P., Moshrefi,M., Paclet,C.M., Park,S., Sequeira,A., Sethi,H.,

Shir,E., Svirskaas,R.R., Weinburg,T. and Ceiniker,S.E.

Direct Submission

Submitted (01-MAR-1993) Berkeley Drosophila Genome Project,

University of California Berkeley, 539 Life Sciences Addition, 3250,

Berkeley, CA 94720, USA

Sequence submitted by:

Berkeley Drosophila Genome Project

University of California Berkeley

Berkeley, CA 94720

For further information about this sequence including its location

and relationship to other sequences, please visit our sequence

archive web site <http://fruitfly.berkeley.edu/sequence/> or send

e-mail to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

source

1..3624

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:7227"

/clone="Lb21405"

454..3093

/codon\_start=

/product="unknown"

/protein\_id="AAD34757.1"

/db\_xref="GI:4972724"

/translation="MAINQGYKKSICAIIDEVDVRSFKLHNSQDSBIGKQKQLOQIEGS

ETAPUTINQCKJHMDGSHLPKDLAAQALNEVAAGGAGGRTPAAITPNSHPSA

NGLRALLPGRSFDALMSLRKRRLIYVTTACICALLL:ILLLAFWEPVFFYLR

PLCDKEGVSSRQLLWANTS KSPCHETYEACNFASDYANHEYPV:KRGEMNYET

YNEVOELNELNRFIAJL:PNSEQSVSTSVLSYSCRETDLVKOSDILLKRA:NAV

YMQAFRDSNRLQSWERYKVLVLCAGHG:FEYVAVTVENGFSKPHDVITL:DEGELG

/PDKYFYGDADDEEVVRSYKLLLRDPFALNMGVSRADLFADDFHYERRIVNH:EDA

KADARPKNLRLADL:KIKAPSUPILSSJQAIFPK:K:TEDEV:VLRDPEVMAHSV

LLSTSDSKPNNFVMSLARKV:PHLSKRYRLTAETFDHALYGR:YASYPRMLICSKV

RMWFFAYDA:CCQPPRQCFETM:SKSRGLNKTQNEELLKMPFSLRNQKDS:EQV

KWLEPAQAQF:CKKLEMLQFG:PDEVLCQPTV:AOYVNELI:LNKVFVEHLE:WT

FERSMEKLGPLA:LQVIVSEMVRDNPQAIAYSNKLNMLVSKVLSSSKAVDYVP

VAVNFARIGTDL:ELALDNSTFELLQNTOSTDI:SEAAPE:RYACQPNVNC:ARAGQPPR

LYHE:NEUSSNALKSFHTLSAARTAAALSNFVNAIDAGSAIQSGIGDQANTYBALG

LTRRLRFGSRSENEELTUSYMQOHCSTTIADKYARIKPLVESQIAESYLEFNATW

CHIQFLPSTNCAAAEATCSN:L"

BASE COUNT 995 a 911 c 884 g 834 t

ORIGIN

Query Match 7.2%; Score 19; DB 3; Length 3624;

Best Local Similarity 100.0%; Pred. No. 138-02;

Matches 19; Conservative C; Mismatches 0; Indels 0; Gaps C;

QY 138 GATCACCTGGACGAGGCG 156

|||||:|||||:|||||

Db 1428 GATCACCTGGACGAGGCG 1446

RESULT 4

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Pseudomonas aeruginosa genes for cytochrome c and cytochrome b

subunits of nitric oxide reductase, complete cds.

3773 bp DNA linear BCT 28 FEB-1999

D38133.1

GI:537365

ANR-binding motif; cytochrome b; cytochrome c; denitrification;

nitric oxide reductase.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (sites)

Arai,H., Igarashi,Y. and Kodama,T.

Structure and ANR-dependent transcription of the ntr genes for

denitrification from Pseudomonas aeruginosa

Biochim. Biotechnol. Biochem. 58 (7), 1286-1291 (1994)

94362287

PUBMED

REFERENCE

2 (bases 1 to 3773)

Arai,H., Igarashi,Y. and Kodama,T.

The structural genes for nitric oxide reductase from Pseudomonas

aeruginosa

Biochim. Biophys. Acta 1261 (2), 279-294 (1993)

95226457

PUBMED

REFERENCE

3 (bases 1 to 3773)

Kodama,T.

Direct Submission

Submitted (02-SEP-1994) Tohru Kodama, Tokyo University, Department

of Biotechnology, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan

(Tel:03-3812-2111)(ex.5142), Fax:03-5684-5196

Submitted (02-Sep-1994) to DDBJ by:

Tohru Kodama

Department of Biotechnology

The University of Tokyo

1-1-1 Yayoi, Bunkyo-ku

Tokyo 113

Japan

Phone: 03-3812-2111 X5142

Fax: 03-5684-5196.

Location/Qualifiers

FEATURES

source	1...3773	Query Match	7.2%; Score 19; DB 1; Length 3773;
	/organism="Pseudomonas aeruginosa"	Best local similarity	100.0%; Pred. No. 1.3e+02;
	/mol_type="genomic DNA"	Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
protein_bird	/strain="PAO1"		
	/db_xref="taxon:287"		
	19...32		
	/note="ANR box"		
	/citation=[1]		
	/citation=[2]		
	/bound_moiety="ANR protein"		
	/function="activation of anaerobic gene expression"		
RBS	80...84		
gene	/gene="nrcC"		
CDS	90...53C		
	/gene="nrcC"		
	/EC number="1.7.99.7"		
	/note="bp270-284 (AA61-64): heme c binding motif of the nrcC gene product."		
	/citation=[2]		
	/codon_start=1		
	/transl_table=1		
	/product="cytochrome c subunit of nitric oxide reductase"		
	/protein_id="BA07329.1"		
	/db_xref="GI:537367"		
	/translation="MSEFTKGMARNYFGSSVFILLFALTYHTKTLPERNEAA		
	XSAAVRLKWNQNCVGHLLGEGAFAPELGVNVRGGEGNIFLQAWNLQ		
	PLNVGRRMPQFHLSEGVDDLAFLKWKSSKIDTQWPNKEG"		
RBS	519...526		
gene	/gene="nrcC"		
CDS	530...1930		
	/gene="nrcB"		
	530...1930		
	/gene="nrcB"		
	/EC number="1.7.99.7"		
	/citation=[2]		
	/codon_start=1		
	/transl_table=1		
	/product="cytochrome b subunit of nitric oxide reductase"		
	/protein_id="BA07330.1"		
	/db_xref="GI:537367"		
	/translation="MMSNGSLKFSACAVKPYFVFALLFVGQILFGLNGLYVVG		
	DFLEPAIFENVARYVHTLILVLLFGFYGAAYLVPEESCELYSKLAWILFWFA		
	AAGVTLILGYSLVYAGLGLTGLNELMPTGYREFLEOPTISKAGIVLALGFLNVM		
	TVLRGRKTAISLVMTGLIGALLFLFSFYNENLTRODKFVWVHLNVEGFWELJM		
	GAILAFVLKVTGDSREVIKWLVIILALISLIGTHHYFPIGVPGVWALGVSF		
	SALFELPFVFLFAFNTINRRRRYDPRNAVALWAGTTVNAFLGAGWGFVETAP		
	VNYVTHGTQLAAAGHMAFYGAYAMVNTIISYAPRARGIGEAADRNSQVLEWGF		
	LNTVMVFITLFLSAGVLQVWLQRMPPADGAAMTFMATQDLA-FYWLREGAGYVPLI		
	GI:VALLSPRRKAAA"		
	1932...3770		
	/function="unknown"		
	/note="ORF6"		
	/citation=[2]		
	/codon_start=1		
	/transl_table=1		
	/product="hypothetical protein"		
	/protein_id="BA07331.1"		
	/db_xref="GI:537368"		
	/translation="MAFAVELEEWGAWHRRPITTHASGEFEARAVLTESVRRPGLM		
	FRALGAPGVLEATPARPLJLRRTWQJRVAGTCEQAFVSWNGSRLPESLAVYPO		
	AELNRELYWALLAASAPLHWAQCNQWARQLLDYAPLRVYARLVAAHLRQRP		
	SLDGPAAAELEALRALAPGVSFFPRVAPVPLWLYPGERWTPSAAGG		
	EEAAGAKROVARGAKPAERIEERNENLILFRLESLSASHEHLADRCDDDD		
	DPDGAVEDLVLISRCQTKGGJLDLDPAADVDLP.LPGC.KLEWQVQCR		
	LLSDHVLQPNRPRGASLAPLHLEKTLHLRQFACLRQGRQLQQCGSDSDGS		
	AWLDQVRRRGSTQPLFUEQRPRLRLACLLADLSMSFEYLDQQRVDS.VV		
	SULLFGEALQALGDPLFGLYGSVRQVRQVRLKDFDEGYGVRGV.LALSPTYT		
	RMGAAIRASQVLPQKQKRELLILLSDCKNDLDRVEGRYGIETPQAVTEARSQGL		
	VPTCTTCKEADYLPVLFADGDFALVERAQLPEELQLYRLR"		
BASE CCJNT	560 a 1117 C 1253 G 633 T		
ORIGIN			



FEATURES  
source

```

location/Qualifiers
1..10711
/organism="Pseudomonas aeruginosa PAO1"
/mol_type="genomic DNA"
/strain="PAO1"
/db_xref="taxon:208954"
139..1921
/gene="nirQ"
/locus_tag="PA0520"
139..1921
/gene="nirQ"
/locus_tag="PA0520"
/locus_tag="PA0520"
/locus_tag="PA0520"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 1"
/codon_start=1
/transl_table=1
/product="regulatory protein NirQ"
/protein_id="AAG03909.1"
/db_xref="GI:9946386"
1921..2278
/translation="NRDAPYVATGHEIEVFBEAWRGVLLKQPTGCGKTFVQY XARLELPVXVACHEDLGAADLGRHLIAGDQWGDGTRAVRGGCYDEUVE ARQDTTVAHPLADRDRELYERTGELQAPPSMLVSNPGYNLLKPKSTPR FVALRFPYPAQQAARCLVSGCAETLAQRLVCGQADRLRLEOHDEEVASTRLIF AARLIGDYPDREACRVALAEPLSDDEATVAALMDIVDLHVA"
956..1483
/gene="nirQ"
/locus_tag="PA0521"
/note="synonym: nirQ"
956..1483
/locus_tag="PA0521"
/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
/codon_start=1
/transl_table=1
/product="probable cytochrome c oxidase subunit"
/protein_id="AAG03910.1"
/db_xref="GI:9946387"
1483..1748
/translation="WKSFLAEVSVESLILAFAGAAKQPELFRAGREGLDSRWLL LVALSSCWAAAVNAVRGRPRQSGNLAAALJLVGLKHUYGRLLADGGL IEDTFFTYWCLTGFEFLVILGLLIGWLAQACWRRCVTPACSGLESGLVYWHW DLVWLLFPLVYVLC"
1491..1748
/locus_tag="PA0522"
/note="synonym: nirP"
1491..1748
/locus_tag="PA0522"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG03911.1"
/db_xref="GI:9946388"
1748..2278
/translation="WRTITLCWJALLAVTGVLLGGAGDSPWLLAAVACAVAKGWL IGERFMSLAHALWRRLLLAMPLLMALAVGAALYLARMNN"
1838..2278
/gene="nirC"
/locus_tag="PA0523"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane-associated; Subcellular localization confidence: Class 1; Alt protein name: cytochrome c subunit of nitric oxide reductase"
/codon_start=1
/transl_table=1
/product="nitric-oxide reductase subunit C"

```

```

/protein_id="AAG03912.1"
/db_xref="GI:9946389"
/translation="XSEITFKGMARNYFGGSVFLLFLALTHITEKILSPRTNEAA MSAAYVRKLQWEQNCVGCHTLLGEGAYFAPELGNVYGRKGEGEFTFQJAXNITC PLAVFGRAPVQPHSEQVEDLAEFJKMSSKIDNTQWPKKRG"
2278..3678
/gene="nirB"
/locus_tag="PA0524"
2278..3678
/gene="nirB"
/locus_tag="PA0524"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1; Alt protein name: cytochrome B subunit of nitric oxide reductase"
/codon_start=1
/transl_table=1
/product="nitric-oxide reductase subunit B"
/protein_id="AAG03913.1"
/db_xref="GI:9946390"
3680..5518
/translation="WXSFNGSKPASQARVAPYVFALLIFVQILFGLINGLQYVVG DFFPAIDFNARMVHTNLLJVLWDFGMAAYLVPEESDCEISPKLAWILFWFA AAGVITILGILLVPYAGIARLTGNEJTMGRETEQETISKAGIVLALGFLENVGM TVRGRKTAISWLVGTGILGALLFPSPFNEMTRCKFYWWVWVHLWVQVWGLSILM GALLAPVVKITGVREVIEWEKLVIYAVALISGIIIGTHHYFVGVPGYVWLGSVF SALLPLPPEAVLFAFNINRRRDYVNAVALKMGCTYVAGLGAQWGFEMHTLAP VNYTHGQTLTAHCHMAFYGAYXVMTIISYAPPLRLEGGEAMQNSQVLEMMGFV LMTVAMVETLIFSANGVQWLQRYPADGAATPMATQDQAIFYWUREGAGVVFLL SLVAYLLSFRGKAAA"
3680..5518
/locus_tag="PA0525"
/note="synonym: nirD"
3680..5518
/locus_tag="PA0525"
/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
/codon_start=1
/transl_table=1
/product="probable dinitrification protein NirD"
/protein_id="AAG03914.1"
/db_xref="GI:9946391"
5518..5791
/translation="WAPAVELEEWGAHSHFITHASGEFEAAVLTESMRPGLM FRALGGAPVVALEATASRLLRRTWLORVAGTCEQAPVSWPNCDSLRLPSLAVPC BELRRELYKWLALLAASAGPLRHQAQDNQVRAQLDAYPAIRPRYARLVAAHLRQRP SLDDCPAADALETALRAJAEPSVERFRVERAPMPVPLWLPGERWTPSAAEDG EAAAGAGKQVARGARKAEIRTEERNRNLFLRLESLSKSHALDRCSDED DPDGAARVADLDYLSLQRTQKGGRLDLDLPADYDLPGLGKLPLWYDVRQQR LLSCHVLQPMRPRGATSAIPAHLEKTLALRRQFACLRDGRMLRQOPQODEIDLQ AWLDFCVRPRGGSTQPGLEORPRRDACLALLADLMSTEAVLDQRVRVDSIVQ SLLLEGALCALGDFEALYGVSSVRCQVWVLKPDDEGYGGEVGRVJALSQYV RMGAIIRASQVLGQCPKRRLLLSLSDGKPNLDRYEGYGIETRAVTEARSQGL VPFCITIDKAAADYLPYJFGADGFALVERAGOLPERLIQYRRUR"
5791..5791
/locus_tag="PA0526"
/complement(5525..5791)
5791..5791
/locus_tag="PA0526"
/complement(5525..5791)
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1
/product="hypothetical protein"

```

Query Match 7.2%; Score 19; DB 1; Length 10711  
 Best Local Similarity 100.0%; Pred.No. 1.3e+02  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 124 CGTGGTGGTGAAGATGAC 144  
 |||||

DB 2949 CGTGGTGGTGAAGATGAC 2969

## RESULT 8

AE012381

LOCUS

AE012381 11004 bp DNA linear BCT 29-MAY-2002  
 Xanthomonas axonopodis pv. citri str. 306, section 459 of 469 of  
 the complete genome.

ACCESSION

AE012381

VERSION

AE012381.1

KEYWORDS

G:21110719

SOURCE

Xanthomonas axonopodis pv. citri str. 306

ORGANISM

Xanthomonas axonopodis pv. citri str. 306

REFERENCE

Xanthomonadaceae; Xanthomonas

1 (bases 1 to 11004)

da Silva, A.C.R., Ferro, J.A., Peirach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Cr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Canargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,

Chambergo, P., Clapina, L.P., Ciccarelli, R.M.B., Coutinho, L.,

Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.J.F., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Cr., A.P., Lemos, E.G.M., Lemos, X.V.P., Locati, E.C., Machado, X.A.,

Madelra, A.X.B.N., Martinez-Rossi, N.M., Martins, E.C., Medeiros, J.,

Menck, C.F.M., Miyaki, C.Y., Xoor, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,

Rossa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitajima, J.P.

2 (bases 1 to 11004)

da Silva, A.C.R., Ferro, J.A., Peirach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Cr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Canargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,

Chambergo, P., Clapina, L.P., Ciccarelli, R.M.B., Coutinho, L.J.,

Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.J.F., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Cr., A.P., Lemos, E.G.M., Lemos, X.V.P., Locati, E.C., Machado, X.A.,

Madelra, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Medeiros, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, C.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,

Rossa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitajima, J.P.

Direct Submission

Submitted (28-NOV-2001): Departamento de Bioquímica, Universidade de

Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,

Brazil

Location/Qualifiers

1. 11004

/organism="Xanthomonas axonopodis pv. citri str. 306"

/mol\_type="genomic DNA"

/strain="306"

/db\_xref="taxon:90486"

/note="pathovar: Citri"

211..679

/gene="XAC4278"

211..879

/gene="XAC4279"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=1

/product="conserved hypothetical protein"

```

/protein_id="AA031113.1"
/db_xref="GI:21110719"
/translation="MTE-SPPVJTTAARVVGGLIEKFAATTDPAYPTVVAQVAAN
QTARSPVNLQGTGVVHIALRQETLGLVPPQFSRRAERYEHLRGLSVLRLTRQVAL:
GLLRGAGTCLGELVARLERLARFENDADMPHLDL:ORGLAAQLPRASGQREDRVA
HL:SGELCVLA:QAAARARARSARSGSDTSCLEARMQALEATVVE:QELAAVQARLD
AAGA"
complement(1156..2271)
/gene="XAC4279"
complement(1156..2271)
/gene="XAC4279"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA031114.1"
/db_xref="GI:21110720"
/translation="NHRILWSTORITAR^HNRVVGSPKDACMSLVVNAVSTLRTPLWP
DLPQAVQHRDHADPALVDH:HGFGVGVNGAGDSQMTQSEVHL:MRHVCRVFEHESFC
VDAGFGLACWASQANAVCLADSVDRPGRVLISGEPALDDAQVPPFDALQR
RAQKSLTAQGIKRVPPSPPPVPGGEARVDAVVGSM:ALFAVALRAEILAGDGT
PPSLDEVTR:PGVAALSPQERAF:QAQAPSAQALNFGMVEYSLAVLQWALGADT
LPEPTALCEVLVAQALDNACAAERP^VARSRLPELLMDLRHLRLHLEWAVRQAGSA
CPVAGIVGVVYERHYALNLLHFDAAEWDSVD^PT"
complement(12527..3081)
/gene="XAC4280"
complement(12527..3081)
/gene="XAC4280"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="acetyltransferase"
/protein_id="AA031115.1"
/db_xref="GI:21110721"
/translation="NMPSCAVTV:QTPRLRLTLPDRDADMLFVNDPQFIAGIN
DRG:ITRQCAHLCEWAQAHARVGFARWAVE^PDAAP^GTLGLLRETLPVPHLG
FALLPAYRQKGVVTEAGRAVLIDYAPVLGLSOLCAIVSPDCAESIRAEIKLQQLQHG
RV:SPETAFVLYADLGFVGRAG"
complement(13159..1513)
/gene="rep"
/note="synonym: XAC4281"
complement(13159..1513)
/gene="rep"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="ATP-dependent DNA helicase"
/protein_id="AA031116.1"
/db_xref="GI:21110722"
/translation="NYGUNPQSAVHCRGRLVLVLAGSGSKTRVIEKIAHLIAG
RYPAP:AAITFNKSKEMREVARP:RGDGAELTICTFHALGKFTQ:EHAAAGL
KRGFS:FSQDAAQ:KDLMHGAKPAIDEDAN^:SRANAGLSPEQMAARSTREX
EAAS:YERYOARLTTFNVDLDCILRLPVLLEAMEDIWGMRRERIGVLLVDECOOT
DCAV:LLKMLAGPRGNFTVCVGDOS:VAMGAPENLQOARDPALAEIKLEONYR
CSNRV:RAANALIAHNEH:KTLMSDADGERIRVWECRDSHEAKVAAEISLGT
KQCVKSPCILFCNFPQSRP:LEKALQ:LRVPHYLTGTGTLERGERVDLLSWRLIV
NPDQAFIRVQSPKREVGAT:ARLAEASAKSVPMVSRASMGALOHUPPRAANG
LSAFTLILRDVREHSSVPAGELVRLADKSGLLNDLRNKSDETGFQRKXLDDELA
EMFEGPRGASQ:AAQALLSRNCKDGGQVRMTMHASKGLEFRYFVIVGCGDG
VLPHEVLEGNLOERELLVYGT^RAKEQ:WMSHSL:TRKPGEHVRLKPRFFDEIP
AEELQDGDGDFVADAEKKAASAGLAAIQALFD"
5341..5970
/gene="tdk"
/note="synonym: XAC4282"
5341..5970
/gene="tdk"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="tyrosine kinase"

```

```

/protein_id="AAM39117.1"
/db_xref="GI:21110723"
/translation="MAKLYFYYSMAAGKTTTLCSAHNYRBMGMLTLITPKLDHRA
GSGVVASRIGLRADCRIFEDTELCOQVEDIHNDGALHCYLVDEAFQFRAQVWYLS
EVDRLEKFLVLCYGLRDFRGEUFEGSCFLWADGALHEEIKTCHGSKATYTVRVA
HGHAVGQPGVETIGNERYVSRAEFKKIRNGEGRIDPLQIALPLEVA"
gene
6391..7647
/gene="XAC4283"
6391..7647
/gene="XAC4283"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAM39118.1"
/db_xref="GI:21110724"
/translation="MEALXDPPLPALADARVGRSSVPVLSVLTITGMSFAVVA
KVDARWITQVHDAISFGLRDEPLATFDCSVVRGNVAFGASDPTFYDHP
SPRYGESYVSVPIKEDDGSVFTGLCALDPQPRVMDAPLVEXVELLAGLLAACIOAE
CRAPESALSRRAESRSGVAGARLSEEFIGILGHDLRPLQSMHADVMSLQOLN
HRYGSSRLSQRSLRVEELIDCAPRGIEIDMLQDLDGASDLDDALSQVDETR
AAYPNQVLSHVAIDEPVHGDAFLAONFGSLKNVAVQPSLLKAVATEHGRLR
TEVNDGIDPRDLVHASTHVTHTGGRALPVDVLDGHRVAAQIAKAHQGSEIRIISKO
GTGFRVSLACTRPRGRHLVGTGAPT"
complement(7814..9415)
/gene="mdog"
/notes="synonym: XAC4284"
complement(7814..9415)
/gene="mdog"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="periplasmic glucan biosynthesis protein"
/protein_id="AAM39119.1"
/db_xref="GI:21110725"
/translation="MQRHFELKAAALAAALGPTLPQWALAAKAYGVRRLQGPDPD
FAMKKGARLANAPYKSHQKLLPGLLEALNMDQYRIRYQDHAWMGNGKFKQAK
FHLGLYFHTPVHYEIVDGRKAQCLAYDPAADFVGRSLGGKQFQKDLGAGPLNTRK
CTDPEFAFQASVFAVRGEGSGVAGSLAIDTGTGGEPEEPDIAYVLEQAFES
DTAVYGLLDSVAGVRRRAITNGSVLNDISALYPRKALERLIGRGTNYQYGE
NKNRMDWRPEHDTUGLAWTGGGWEWRPJCNPFLHKNFVDENRSGELLQPD
RNDHYDGGVFEKRCLEQVFKSGKGSQVLEVTPTVETFDVIAFWNPQAKPC
FGSLLXGLYLAQCPAFSSPAHCVATRTGIGVIGQKSHSFSAVTFAGSGLA
ALAKDPKAYEAVIQVSRGTETVSAEPLHKLGYRAMELVPVDEGTQQIDIRLR
ANGKPLTETWYQWTPPTGERRRY"
complement(9494..10306)
/gene="XAC4285"
complement(9494..10306)
/gene="XAC4285"
/notes="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM39120.1"
/db_xref="GI:21110726"
/translation="MRPNRSGSLLETVETSPNALRPGQLQADIARTGTWPAATCCQA"

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 216 GTTGGGAGCTGTATGATC 234
      |||||
Db 7332 GTTGGGAGCTGTATGATC 7320

11554 bp DNA linear ECT 23 MAY-2002
AE012540
LOCUS
DEFINITION
Xanthomonas campestris pv. campestris str. ATCC 33913, section 448
of 460 of the complete genome.

```

```

ACCESSION
AE012540 AE009322
VERSION
AE012540.1 GI:2111542f
KEYWORDS
Xanthomonas campestris pv. campestris str. ATCC 13913
Xanthomonas campestris pv. campestris str. ATCC 13913
Bacteria; Proteobacteria; Gammaproteobacteria, Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
REFERENCE
1 (bases 1 to 11554)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Canavan, P., Cardoso, J.,
Chambergo, F., Clapina, P., Cicatelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., E-Dorrry, H., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, X.L., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lenos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madelra, A.V.B.N., Martinez-Rossi, N.Y., Martins, E.C., Meidanis, C.,
Venck, C.F.X., Miyaki, C.Y., Voon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, N.C., Oliveira, V.R., Pereira, C.H.A.,
Rosa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, J.A.F.,
Takita, M.A., Tanura, R.E., Teixeira, E.C., Tezza, R.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002)
JOURNAL
2222145
MEDLINE
1234217
PUBMED
1234217
REFERENCE
2 (bases 1 to 11554)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Canavan, P., Cardoso, J.,
Chambergo, F., Clapina, P., Cicatelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., E-Dorrry, H., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, X.L., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lenos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madelra, A.V.B.N., Martinez-Rossi, N.Y., Martins, E.C., Meidanis, C.,
Venck, C.F.X., Miyaki, C.Y., Voon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, N.C., Oliveira, V.R., Pereira, C.H.A.,
Rosa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, J.A.F.,
Takita, M.A., Tanura, R.E., Teixeira, E.C., Tezza, R.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.
Direct Submission
Submitted (23-NOV-2001) Departamento de Bioquímica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508 900,
Brazil
FEATURES
Location/Qualifiers
1..11554
/organism="Xanthomonas campestris pv. campestris str. ATCC
33913"
/molecule_type="genomic DNA"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
199..1446
/gene="XCC4144"
199..1446
/gene="XCC4144"
/feature="XCC4144"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAM43365.1"
/db_xref="GI:21115427"
/translation="MGOLPSELHADVARVGLSSVPVLTITLTGKREANAVARVTD
ARWTCGVQVQLAFSGQFGLPATTFCGVRNRQVAFSGHAFVTRDHPSPKL
YGFESYSEVFRLDGGSVFGLQADPEPRWDAPLVEKIELLQALIGLQAOIELRAS

```



30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
 RP4-646P11 is from the library RPC1-4 constructed by the group of Pieter de Jong. For further details see <http://www.chcri.org/bacpac/home.htm>  
 VECTOR: pCYPAC2.

#### FEATURES

source  
 1..14616  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP4-646P11"  
 /clone\_11b="RPC1-4"

BASE COUNT 4432 a 2869 c 342 g 4173 t

#### FEATURES

Query Match 7.2%, Score 19; DB 9; Length 14616;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 20 AGCCAGCAGGCCAGCTT 38  
 |||||  
 Db 6477 AGCCAGCAGGCCAGCTT 6495

#### RESULT 11

AC017508 20686 bp DNA linear HTG 09-DEC-1999  
 LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION  
 AC017508  
 ACCESSION  
 QY 20 AGCCAGCAGGCCAGCTT 38  
 |||||  
 Db 6477 AGCCAGCAGGCCAGCTT 6495  
 KEYWORDS  
 HTG; HTGS\_PHASE2.  
 SOURCE  
 Drosophila melanogaster (fruit fly)  
 ORGANISM  
 Eukaryota; Metazoa; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 20686)  
 Adams, M. and Venter, J. C.  
 Direct Submission  
 Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 This sequence was identified as CDM:10211070 by the submitter.  
 For more information on this record e-mail to [f.y@celera.com](mailto:f.y@celera.com).  
 \* NOTE: This is a 'working draft' sequence.  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

#### REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 This sequence was identified as CDM:10211070 by the submitter.  
 For more information on this record e-mail to [f.y@celera.com](mailto:f.y@celera.com).  
 \* NOTE: This is a 'working draft' sequence.  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

#### FEATURES

source  
 1..20686  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 BASE COUNT 6189 a 4224 c 4223 g 6050 t  
 ORIGIN  
 Query Match 7.2%, Score 19; DB 2; Length 25686;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

78 CAGCCTGACAGCAGCTG 96  
 |||||  
 Db 16312 CAGCCTGACAGCAGCTG 16330

#### RESULT 12

AC015098 74433 bp DNA linear HTG 16-NOV-1999  
 LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION

#### ACCESSION

VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

#### REFERENCE

AUTHORS  
 TITLE  
 JOURNAL

#### COMMENT

This sequence was identified as CDM:10214196 by the submitter.  
 For further information on this sequence e-mail to [f.y@celera.com](mailto:f.y@celera.com).  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

#### FEATURES

source  
 1..74433  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"

BASE COUNT 22351 a 15293 c 15492 g 21297 t  
 ORIGIN

Query Match 7.2%, Score 19; DB 2; Length 74433;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

139 GATCACCTGACGAGGCG 156  
 |||||  
 Db 12839 GATCACCTGACGAGGCG 12821

#### RESULT 13

AL596222 107056 bp DNA linear PRI 09-FEB-2002  
 LOCUS Human DNA sequence from clone RP1-1140:8 on chromosome 1, complete  
 DEFINITION  
 AC596222  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 107056)  
 Hopkins, B.  
 Direct Submission  
 Submitted (09-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail: enquiries.  
 enquiries@sanger.ac.uk  
 Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Feb 11, 2002 this sequence version replaced GI:18335166.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

#### REFERENCE

AUTHORS  
 TITLE  
 JOURNAL

#### COMMENT

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: EMBL, EMBL; SW,  
 SWISSPROT; Tr, TrEMBL; Wp, WormPep; information on the WormPep  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence

was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>  
 RP11-114018 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTER: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-114018. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true right end of clone RP11-114018 is at 1070356 in this sequence. The true right end of clone RPI1-323X8 is at 2002 in this sequence.

## FEATURES

source

Location/Qualifiers  
 1..1070356  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-114018"  
 /clone\_lib="RPI1-11.1"  
 32752 a 22325 c 20837 g 31138 t

Query Match 7.2% Score 19 DB 9 Length 1070356;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 20 AGCCGAGCCAGCCCGAGTT 38

Db 59311 AGCCGAGCCAGCCCGAGTT 59913

## RESULT 14

ACC10056/c

LOCUS

DEFINITION

Drosophila melanogaster chromosome 1L/79A3 clone RPI198-25D3, \*\*\*

SEQUENCING IN PROGRESS \*\* 52 unordered pieces.

ACC10056

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Diptera; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 135055)

Munzy,D.M., Adams,C., Bailey,X., Barbara,J., Blankenburg,K.,

Bodota,B., Bouck,C., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,

Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,

David,R., Delgado,C., Deshaize,D., Ding,Y., Domar-Rastid,N.,

Duan-Rocha,S., Dubin,K.H., Fernandez,C., Ferraguto,D.,

Fitzcum-Tansey,J., Franz,P., Ganesh,P., Gore,J.H., Gorrill,L.L.,

Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,

Hollaway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,X.,

Kelly,S., Kondrjewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,

Lichtarge,C., Liu,J., Liu,W., Logan,C., Loeb,J., Lucier,R.,

Martin,R., Martinez,C., Mebeck,M.P., Mei,G., Morgan,M., Morris,S.,

Nash,S., Nelson,A., Nguyen,N., Nguyen,S., Osval,G.,

Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,X.,

Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,

Sher,H., Simon,M., Sparks,A., Stamps,A., Suggang,R., Tabor,P.,

Taylor,T., Vasquez,L., Vinsen,R., Vo,C., Wabba,M., Wattington,S.,

Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,

Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

Direct Submission

Unpublished

REFERENCE

2 (bases 1 to 135055)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

AUTHORS

Submitted (11-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REMARK  
 COMMENT

Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 135055)  
 Worley,K.C.  
 Direct Submission  
 Submitted (29-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 Sequence format modified by NCBI staff  
 On Jan 29, 2003 this sequence version replaced gi:6056152.  
 NOTE: This is a 'working draft' sequence. It currently consists of 52 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

820: contig of 820 bp in length  
 821: gap of 100 bp  
 822: contig of 1072 bp in length  
 1391: gap of 100 bp  
 2393: contig of 1246 bp in length  
 3339: gap of 100 bp  
 3439: contig of 1123 bp in length  
 4562: gap of 100 bp  
 4662: contig of 858 bp in length  
 5519: gap of 100 bp  
 5520: gap of 100 bp  
 5620: contig of 1450 bp in length  
 7070: gap of 100 bp  
 7170: contig of 832 bp in length  
 8002: gap of 100 bp  
 8102: contig of 826 bp in length  
 8328: gap of 100 bp  
 8328: contig of 1034 bp in length  
 10362: gap of 100 bp  
 10362: contig of 714 bp in length  
 10876: gap of 100 bp  
 10976: contig of 581 bp in length  
 11557: gap of 100 bp  
 11557: contig of 935 bp in length  
 11557: gap of 100 bp  
 12592: gap of 100 bp  
 12592: contig of 846 bp in length  
 13438: gap of 100 bp  
 13438: contig of 1834 bp in length  
 13772: gap of 100 bp  
 13772: contig of 1876 bp in length  
 17148: gap of 100 bp  
 17148: contig of 1556 bp in length  
 18553: gap of 100 bp  
 18553: contig of 1025 bp in length  
 19578: gap of 100 bp  
 19578: contig of 1441 bp in length  
 21420: gap of 100 bp  
 21420: contig of 1115 bp in length  
 22334: gap of 100 bp  
 22334: contig of 1501 bp in length  
 24335: gap of 100 bp  
 24335: contig of 1547 bp in length  
 26382: gap of 100 bp  
 26382: contig of 1234 bp in length  
 27416: gap of 100 bp  
 27416: contig of 1750 bp in length  
 29366: gap of 100 bp  
 29366: contig of 1882 bp in length  
 31348: gap of 100 bp  
 31348: contig of 1350 bp in length  
 32698: gap of 100 bp  
 32698: contig of 1065 bp in length  
 33864: gap of 100 bp  
 33864: contig of 2280 bp in length  
 36343: gap of 100 bp  
 36343: contig of 1305 bp in length  
 37649: gap of 100 bp



```

* 12592 13437: contig of 846 bp in length
* 13438 13537: gap of 100 bp
* 12533 13771: contig of 1834 bp in length
* 15372 15471: gap of 100 bp
* 15472 17147: contig of 1676 bp in length
* 17148 17247: gap of 100 bp
* 17248 18753: contig of 1506 bp in length
* 18754 18853: gap of 100 bp
* 18854 18878: contig of 1025 bp in length
* 18879 18978: gap of 100 bp
* 19979 21419: contig of 1441 bp in length
* 21420 21519: gap of 100 bp
* 21520 22834: contig of 1115 bp in length
* 22835 22934: gap of 100 bp
* 22935 24335: contig of 1501 bp in length
* 24336 24535: gap of 100 bp
* 24536 26082: contig of 1547 bp in length
* 26083 26182: gap of 100 bp
* 26183 27416: contig of 1234 bp in length
* 27417 27516: gap of 100 bp
* 27517 29266: contig of 1750 bp in length
* 29267 29366: gap of 100 bp
* 29367 31248: contig of 1882 bp in length
* 31249 31348: gap of 100 bp
* 31349 32698: contig of 1350 bp in length
* 32699 32798: gap of 100 bp
* 32799 33863: contig of 1065 bp in length
* 33864 33963: gap of 100 bp
* 33964 36243: contig of 2280 bp in length
* 36244 36343: gap of 100 bp
* 36344 37648: contig of 1305 bp in length
* 37649 37748: gap of 100 bp
* 37749 39440: contig of 1692 bp in length
* 39441 39540: gap of 100 bp
* 39541 41760: contig of 2220 bp in length
* 41761 41860: gap of 100 bp
* 41861 43610: contig of 1750 bp in length
* 43611 43710: gap of 100 bp
* 43711 46417: contig of 2707 bp in length
* 46418 46517: gap of 100 bp
* 46518 48732: contig of 2215 bp in length
* 48733 48832: gap of 100 bp
* 48833 51428: contig of 2596 bp in length
* 51429 51528: gap of 100 bp
* 51529 51987: contig of 2459 bp in length
* 51988 54087: gap of 100 bp
* 54088 56621: contig of 2534 bp in length
* 56622 56721: gap of 100 bp
* 56722 59477: contig of 2756 bp in length
* 59478 59577: gap of 100 bp
* 59578 62712: contig of 3135 bp in length
* 62713 62812: gap of 100 bp
* 62813 64882: contig of 2070 bp in length
* 64883 64982: gap of 100 bp
* 64983 69159: contig of 4177 bp in length
* 69160 69259: gap of 100 bp
* 69260 72255: contig of 2396 bp in length
* 72256 72355: gap of 100 bp
* 72356 75159: contig of 3004 bp in length
* 75160 75459: gap of 100 bp
* 75460 78631: contig of 3172 bp in length
* 78632 78731: gap of 100 bp
* 78732 81527: contig of 2796 bp in length
* 81528 81627: gap of 100 bp
* 81628 84656: contig of 3029 bp in length
* 84657 84756: gap of 100 bp
* 84757 89863: contig of 5107 bp in length
* 89864 89963: gap of 100 bp
* 89964 94990: contig of 5027 bp in length
* 94991 95090: gap of 100 bp
* 95091 100921: contig of 5902 bp in length
* 100922 100993: gap of 100 bp
* 100994 107412: contig of 6320 bp in length

```

```

* 107413 107512: gap of 100 bp
* 107513 115449: contig of 7937 bp in length
* 115450 115449: gap of 100 bp
* 115550 123161: contig of 7612 bp in length
* 123162 123261: gap of 100 bp
* 123262 135055: contig of 11794 bp in length
FEATURES
    location/Qualifiers
        ..135055
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /chromosome="3L:79A3"
        /clone="RP1193-25D3"

```

```

BASE COUNT  35760 a 28456 c 29103 g 36584 t      150 others
ORIGIN

```

```

Query Match      7.2%  Score 19, DB 2: Length 135055,
Best local similarity 100.0%, Pred. No. 1: 2e+02,
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      78 CAGCCTGAACGAGCAGCTG 96
      ||||| ||| ||| |||
Db      82577 CAGCCTGAACGAGCAGCTG 82559

```

```

Search completed: November 7, 2003, 14:25:12
Job time: 1517 secs

```